**SUPPLEMENTARY MATERIALS**

**Comparative validation of breast cancer risk models and projections for future risk stratification**

Parichoy Pal Choudhury,1\* Amber N. Wilcox,2\* Mark N. Brook,3 Yan Zhang,1 Thomas Ahearn,2 Nick Orr,4 Penny Coulson,3 Minouk J. Schoemaker,3 Michael E. Jones,3 Mitchell H. Gail,2 Anthony J. Swerdlow,3,5 Nilanjan Chatterjee,\*\*1,6 Montserrat Garcia-Closas\*\*2

**Affiliations:**

1Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, USA

2Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, USA

3Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK

4Centre for Cancer Research and Cell Biology, Queen’s University Belfast, Belfast UK

5Division of Breast Cancer Research, The Institute of Cancer Research, London, UK

6Department of Oncology, School of Medicine, Johns Hopkins University, USA

The authors have no conflicts of interest to disclose.

\*These authors contributed equally to the manuscript.

\*\*These authors contributed equally to the manuscript.

1. **Cohort Populations for Validation**

1.1 Definition of Follow-Up

The Generations Study (GS) subjects received their first and second follow-up questionnaires approximately 2.5 years (median=2.53, IQR=2.48, 2.59) and 6 years (median=6.13, IQR=5.91, 6.46) after study entry, respectively. In addition to risk factors, the questionnaires sought information on breast cancer diagnosis.1 Breast cancer reports were confirmed by cancer registry, hospital or pathology records. Follow-up was defined from the date of study entry to the date the latest of the follow-up questionnaires was due. A small fraction of the participants (2.9%) were lost to questionnaire follow-up and most of them (~97%) agreed to be flagged with the National Health Service Central Registers to determine breast cancer and vital status. These participants were censored at June 30, 2012 because (at the time of data preparation) national cancer incidence data were incomplete after that date; the remaining participants were censored at loss to follow-up.

In the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial cohort (PLCO) subjects were followed up annually by the recruitment centers to ascertain cancer diagnosis status until trial year 13 or until December 31st 2009, whichever came first. Analyses included women from both the control and intervention arms of PLCO. Sensitivity analyses restricted to the control arm yield similar results, therefore, we only provide full cohort analyses.

1.2 Age Range of the Validation Cohorts

Subjects less than 35 years of age at entry were excluded because this is the minimum age for the BCRAT model and subjects greater than 75 years of age at entry were excluded so that five-year risk predictions are not made beyond 80 years of age.

1.3 Risk Factors in the Validation Cohorts

There are several risk factors in the IBIS model for which the two validation cohorts in this analysis (GS and PLCO) did not have data: benign breast disease (BBD) pathology, Ashkenazi inheritance, personal BRCA 1/2 status and status of relatives, bilateral breast cancer (status and age), number of nieces, and data on cousins. Additionally, the GS did not have information on personal history of ovarian cancer, as well as ovarian cancer status and breast or ovarian cancer age among aunts. PLCO was additionally missing information on (hormone replacement therapy) HRT use, as well as data on grandmothers and aunts.

Due to certain limitations of the PLCO questionnaires, some assumptions were made when performing the IBIS model analysis in this cohort. For subjects with more than one relative of a particular type (i.e., sister or daughter), the questionnaire provided a summary count of the total number of each type of cancer (breast and ovarian) among the relatives. In such situations, we assumed that there were multiple relatives, each with a cancer, rather than multiple cancers in a single relative. In addition, the baseline questionnaire, completed between 1993-2001, provided an aggregate count for number of sisters (not distinguishing between half and full). Therefore, number of full sisters was ascertained from the supplemental questionnaire, completed between 2006-2008.

For the BCRAT model, both the GS and PLCO are missing data on history of atypical hyperplasia. PLCO is missing data on number of breast biopsies, so we assume that subjects with a history of BBD have had one breast biopsy, otherwise we assume no breast biopsies.

PLCO is missing data on HRT type among current users for the iCARE-Lit model and HRT type among ever users for the iCARE-BPC3 model.

**2.0 Breast Cancer Risk Models**

Below, we describe the breast cancer risk models evaluated in this analysis: iCARE-Lit and iCARE-BPC3 models and two established models (BCRAT and IBIS). **Supplementary Table 1** shows the risk factors included in each model.

2.1 iCARE-Based Risk Models

We developed the iCARE-Lit model by combining estimates of relative risk parameters obtained from a literature review. The iCARE-Lit model has two sets of risk estimates and risk factor distributions: one for subjects less than 50 years of age and another for subjects 50 years of age or greater (**Supplementary Table 2**). This age stratification accounts for the fact that the relative risks for some risk factors (e.g., BMI and family history) are modified by age or menopausal status. This also accounts for the age-dependent distribution of some risk factors (e.g., parity and oral contraceptive use).

There are several differences in the specification of the iCARE-Lit model for women less than 50 years of age compared to the iCARE-Lit model for women 50 years of age or greater. iCARE-Lit (<50) includes two oral contraceptive (OC) use variables (never versus ever, current versus former or never) in order to be more flexible for validation cohorts that only have data on never/ever use. When there is no information on whether the ever OC user is a current or former user, this information is imputed using the reference dataset. The iCARE-Lit (≥50) model includes only never/ever OC use, assuming that all ever users are former users. iCARE-Lit (≥50) includes age at menopause and HRT use, while the iCARE-Lit (<50) model does not include these risk factors. Additionally, the relative risks for BBD, BMI, and family history of breast cancer vary between the iCARE-Lit (<50) and iCARE-Lit (≥50) models.

We previously developed the iCARE-BPC3 model for predicting absolute risk of breast cancer for White women in the US 50 years of age or greater.2 The model was derived based on relative risk parameters estimated from cohort studies participating in the Breast and Prostate Cancer Cohort Consortium (BPC3) (17,171 cases and 19,862 controls).

To predict risk for the UK-based cohort (GS), in both models we used incidence rates for invasive breast cancer and competing mortality rates from the UK Office for National Statistics (2006).3,4 Similarly, to predict risk for the US-based cohort (PLCO), we used incidence rates for invasive breast cancer from the US National Cancer Institute-Surveillance, Epidemiology, and End Results Program (NCI-SEER) (2008-2012)5 and competing mortality rates from the Center for Disease Control (CDC) WONDER database (2008-2012).6 We used information from various population-based surveys for estimating distribution of risk factors in both models. **Supplementary Table 2** and **Supplementary Table 3** describe the sources of information for the distribution of risk factors that are used to represent the UK and US populations.

2.2 BCRAT Risk Model

The Breast Cancer Risk Assessment Tool (BCRAT), aka “Gail model”, was first developed in 1989 using data from a case-control study (2,852 cases and 3,146 controls) nested in the Breast Cancer Detection and Demonstration Project (BCDDP), a cohort of women undergoing annual screening mammogram in the US.7 The original model was subsequently modified in 1999 to project risk of developing invasive breast cancer using age-specific incidence rates from NCI-SEER and attributable risk estimates based on Cancer and Steroid Hormone (CASH) Study (<https://seer.cancer.gov/archive/studies/epidemiology/study14.html>), and the available algorithm used in this report was last updated in 2011 (<http://www.cancer.gov/bcrisktool/>).8 Relative risk scores and absolute risk estimates were obtained from the BCRAT Macro Version 3.0 (May 2011). Predictions for both validation cohorts were calculated using the BCRAT default incidence rates (i.e., SEER 1983-87 incidence rates).

2.3 IBIS Risk Model

The International Breast Cancer Intervention Study model (IBIS version 8, aka“Tyrer–Cuzick Model”, <http://www.ems-trials.org/riskevaluator/>) uses family history information to compute a woman’s likelihood of carrying genes predisposing to breast cancer (in particular, BRCA1, BRCA2 and an additional low penetrance gene).9 The likelihood of carrying these genes is then used in conjunction with information on classical risk factors to estimate the probability of developing breast cancer over a specified period of time. We calculated expected five-year risk for both validation cohorts using software provided by the developers, which uses 2008-2010 breast cancer incidence and competing mortality rates obtained from Cancer Research UK.

**3.0 Derivation of the Reference Datasets**

In order to project the distribution of absolute risk for the US and UK populations, we generated risk factor distributions that are representative of each population. These risk factor distributions were each coded according to the iCARE-BPC3 and iCARE-Lit models. Risk factor distributions are specified separately for women less than 50 years of age and women 50 years of age or older. These age categories are used as surrogates for menopausal status: we assume that women less than 50 years of age are premenopausal and never users of HRT, and women 50 years of age or older are postmenopausal and not current users of OC.

3.1 US Reference Dataset

The US reference dataset is adapted from the reference dataset developed by Maas et al.2 The majority of the risk factors were derived from the 2008, 2010, and 2012 National Health and Nutrition Examination Survey (NHANES). An imputation model was used to simulate the distribution of alcohol intake based on the distribution among the controls in the Women’s Health Initiative (WHI) study. The imputation model was conditional on all variables in the iCARE-BPC3 model with significant associations with alcohol intake. The general strategy for imputation was to transform the variable to be normally distributed and then model the transformed variable conditional on other variables. At first the model was considered based on the regression of the transformed outcome based on each predictor separately to explore if there was a statistically significant association. In the second step, a joint model was fit that included all the predictors that were significantly associated with the transformed alcohol intake variable when considered individually. In this joint model, some of the predictors did not remain statistically significant and they were subsequently dropped one at a time starting with the one with the largest p-value. At each step likelihood ratio tests were used to compare each model with the reduced model after dropping the variable. After this process, the final model was used to impute the transformed outcome and that was subsequently back-transformed to get the actual outcome. A second model was used to determine whether the subject was a non-drinker. For subjects who were predicted to be non-drinkers, their alcohol consumption was set to zero, over-writing the previous imputation process. In the final step, the alcohol intake for each referent subject was imputed as an average predicted value plus a sampled value from the model residuals. The distribution of the imputed variable was close to the empirical data distribution. The distributions of the remaining risk factors were derived from the 2010 National Health Interview Survey (NHIS), the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial, and published literature.

3.2 UK Reference Dataset

The risk factor distribution representative of the general UK population was simulated based on population-based surveys, cohort studies, and the published literature. The majority of variables were derived from the Health Survey for England (HSE) of 2005-2006. To account for missing data on OC and HRT status, the joint distribution was simulated for women 50 years of age or greater based on the conditional probabilities among those with data on these variables. In doing so, we are assuming that the status of HRT and OC use are missing at random.

In order to account for the correlation between parity and age at first birth, the joint distribution was derived from the Cohort Fertility Tables, England and Wales. In order to capture temporal affects in parity and age at first birth, the frequencies of these variables were averaged among women born between 1930 and 1955 to represent those aged 50 years or greater in 2006, and among women born between 1960 and 1975 to represent those less than 50 years of age in 2006. The contribution of each age interval from the fertility tables to the overall averages was weighted based on the age distribution of participants in the HSE surveys.

Multiple data sources were used to account for the correlation between body mass index (BMI) and age at menarche. Since the distribution of BMI in the GS would not be representative of the general UK population, the joint distribution of these two variables was simulated from the distribution of BMI in HSE, the distribution of age at menarche in the GS, and covariance between BMI and age at menarche in the GS.

The distributions of the remaining risk factors were derived from published literature.

**4.0 Validation of Breast Cancer Risk Models**

For each subject in the validation cohorts, person-time was accrued from the time of recruitment until the time of last contact or linkage with the subject. The follow-up time for each person for five-year risk prediction was defined from time of entry to five years after entry, time of last contact or linkage to cancer or death registries, whichever came first. The absolute risk of a woman developing breast cancer over the follow-up period was calculated accounting for competing risk due to death from other causes. From each risk prediction model (except IBIS), we obtained the relative risk score (i.e., linear predictor associated with the risk factors except age) and the expected absolute risk over five years since study entry for each subject in the validation cohorts. As the IBIS software does not provide relative risk scores, they were approximated as the ratio of the predicted absolute risk over one year since study entry and the incidence rate of invasive breast cancer at age of study entry.

The study subjects were classified into categories of low to high risk based on (a) deciles of the expected absolute risk, (b) deciles of the relative risk score. Within each risk category, the observed proportion of cases and the average of the expected absolute risks over five years were computed after adjusting for observed follow-up.10 The 95% Wald-based confidence intervals for the observed proportion of cases were computed using an influence function-based variance estimator.10 The expected-to-observed ratio (E/O) was defined (overall and within each risk category) as the ratio of the average of expected risk (overall and within each risk category) and the observed proportion of cases (overall and within each risk category). We compute a 95% Wald-based confidence interval for the expected-to-observed ratio using its asymptotic variance estimator. When the 95% confidence interval (overall or within a risk category) includes 1, we conclude that the model is well calibrated (overall or within that category). Other possibilities include, overestimation of risk, where this ratio is greater than 1 and underestimation of risk, where this ratio is less than 1. In both the scenarios, the 95% confidence interval excludes 1.

For each category, we also computed the observed and expected relative risks. Within each risk category, the observed relative risk was defined as the ratio of the proportion of cases in that category and the proportion of cases in the study. The expected relative risk in each category was defined as the ratio of average five-year absolute risk in that category and the average five-year absolute risk in the validation study. The 95% Wald-based confidence intervals for observed relative risk were computed using the asymptotic variance estimator.10

We evaluated the overall discriminatory ability of each model using the Area Under the Curve (AUC), which was estimated as the empirical proportion of case-control pairs in which the absolute risk (or relative risk score for age-adjusted AUC) for a case was higher than the absolute risk (or relative risk score for age-adjusted AUC) for a control. We also computed the 95% Wald-based confidence interval for AUC using the asymptotic variance formula derived by DeLong et al.11

**5.0 Breast Cancer Risk Projection**

As a first step, we simulated current ages based on the distribution of ages in the 2016 US Census estimates for women 50-70 years in the US. In the next step, we estimate the five-year absolute risk and relative risk scores using the iCARE-BPC3 model and information from simulated current ages, the risk factor distribution from the reference dataset, and 2015 invasive breast cancer incidence rates from SEER.12 We use the standard deviation of the relative risk score evaluated based on the risk factor distribution to calculate theoretical AUCs for the projected risks.13,14 The theoretical AUCs were computed using a normal approximation of the relative risk scores for different models.13-15 When the risk score is defined by a combination of many factors, one can reasonably assume its distribution to follow a normal distribution in the population due to the classic Central Limit Theorem. Thus, one can assume the distribution of risk score in the population can be characterized by it mean and variance . If the disease is rare, the distribution of risk scores in controls can be reasonably approximated by the same normal distribution as that of the population. Now if the risk score is well calibrated in the log-risk scale, i.e , then by simple application of Bayes theorem, one can show that the distribution of risk scores among cases follows a normal distribution with mean and variance Because AUC is simply defined the probability of a randomly selected case will have risk score higher than that of a randomly selected control, i.e. , based on the normal theory one can derive

The mid-2016 population estimates from the US Census Bureau are used to translate the proportion of the population and of future cases identified as exceeding a pre-specified absolute risk threshold into corresponding numbers of subjects. We have considered two high-risk thresholds corresponding to five-year absolute risk of breast cancer crossing 3% and 6%, respectively. The 3% five-year risk threshold is used for recommending risk-lowering drugs (e.g., tamoxifen or raloxifene) by the United States Preventive Services Task Force (USPSTF).16The 6% five-year risk thresholdis used for very high risk in the WISDOM trial and corresponds to the average risk of BRCA carriers.17 At the lower risk, we have used thresholds corresponding to the absolute risk of breast cancer below 1.13% and 0.6%. The 1.13% five-year risk threshold corresponds to the average five-year risk for a 50-year old US woman and the 0.6% five-year risk threshold corresponds to the average five-year risk for a 40-year old US woman, based on the age-specific rates of invasive breast cancer from SEER. We have used these thresholds because these are the recommended ages for starting mammographic screening under different guidelines (i.e., USPSTF, American Cancer Society, the NCCN Clinical Practice Guidelines in Oncology, and the American College of Obstetricians and Gynecologists).16,18-20

The theoretical AUC depicts the amount of overall risk stratification achieved by the various risk factor combinations. For the risk factors for which reference data were not available (e.g., PRS, mammographic density), the relative risk scores were simulated using published estimates of their population level summary statistics (C. Vachon, unpublished data, 2018).21-24 Mammographic density was defined in relation to percent density and the log-odds ratio was adjusted for age and BMI (C. Vachon, unpublished data, 2018). We explored three versions of PRS: (1) the “current” PRS based on 313 SNPs,21,23,24 (2) an “improved” PRS, which could be built by doubling the sample size relative to current study (i.e., approximately 300,000 cases and 300,000 controls) where an underlying optimal p-value threshold (p-value=0.003) are both derived based on theory we developed earlier14 and estimates of underlying effect-size distribution obtained using the GENESIS methodology25 we recently developed; (3) and the “maximum” PRS that could explain heritability for breast cancer associated with log-additive effects of all common variants.25 These calculations were done using estimates of number of underlying susceptibility SNPs and distribution of their effects, which we obtained through application of a novel method for analysis of effect-size distribution underlying GWAS.25 With the overall familial relative risk (FRR) of breast cancer being 2,21 the overall heritability in the frailty scale or log-odds scale is calculated using the formula: = 1.39. The “current” PRS is based on the 313 common susceptibility variants that explain 18.5% of the overall FRR. The heritability explained by the “current” PRS is . The heritability associated with the “improved” PRS (0.37) is calculated using a projection formula14,25 that uses information from the estimated common variant effect size distribution using the summary level statistics and external LD information. The “maximum” PRS based on all the SNPs that can be reliably imputed using OncoArray explain 41% of the overall FRR. The heritability explained by the “maximum” PRS is .21 In the simulations, we assume that the PRS is independent of epidemiologic risk factors given family history. For projections based on risk factor combinations with PRS, family history and other risk factors, we account for the attenuation of the relative risk for family history due to the correlation between PRS and family history.

5.1 Net Benefit for iCARE-BPC3 model

We used decision curves to compare the predicted net benefit of the different risk models for interventions at different high-risk thresholds. We constructed theoretical net benefit curves from the distributions of predicted risk (assuming perfect calibration) from the iCARE-BPC3 model and its extensions after addition of PRS and MD in the US population of non-Hispanic White women aged 50-70 years. We define net benefit for a high-risk decision at the risk threshold by the formula: , where is the number of true positives (i.e., number of women who are above the risk threshold and developed invasive breast cancer within five-years), is the number of false positives (i.e., number of women who are above the risk threshold and did not develop invasive breast cancer within five years) and is the total number of women in the population.26-29 The net benefit can be interpreted as the number of women above the risk-threshold who would develop breast cancer (*TP*) within five years and benefit from an intervention, adjusted for the detrimental effects of intervening on women above the risk-threshold who do not develop breast cancer within five years (*FP*), expressed as a percentage of the total population. The net benefit of the risk models is compared to two clinical alternatives of intervening no one versus intervening all women irrespective of their risk score. These comparisons allow us to put into perspective whether the use of risk models to select which women should be intervened provides any benefit (or harm), compared to intervening on everyone or not intervening at all in the population.

**Supplementary Table 1.** Risk factor distributions in the validation cohorts\*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Breast cancer risk factor** | **GS** | | | | **PLCO** | |
| **Ages <50, N=28,232** | | **Ages ≥50, N=36,642** | | **Ages ≥50, N=48,279** | |
| **Age at baseline, years** |  | |  | |  | |
| Median (range) | 42 | (35-49) | 58 | (50-74) | 61 | (50-75) |
| **Age at menarche, years** |  | |  | |  | |
| ≤11 | 5,600 | (22.4) | 7,775 | (23.3) | 9,612 | (19.9) |
| 12-13 | 12,836 | (51.2) | 16,210 | (48.7) | 26,326 | (54.6) |
| 14-15 | 5,777 | (23.1) | 8,255 | (24.8) | 10,224 | (21.2) |
| ≥16 | 834 | (3.3) | 1,077 | (3.2) | 2,024 | (4.2) |
| Missing | 3,185 |  | 3,325 |  | 93 |  |
| **Parity** |  | |  | |  | |
| Nulliparous | 5,677 | (20.1) | 4,524 | (12.4) | 4,149 | (8.6) |
| 1 birth | 4,155 | (14.7) | 3,710 | (10.1) | 3,270 | (6.8) |
| 2 births | 12,879 | (45.7) | 1,824 | (50.9) | 11,523 | (23.9) |
| 3 births | 5,491 | (19.5) | 9,734 | (26.6) | 29,270 | (60.7) |
| Missing | 30 |  | 50 |  | 67 |  |
| **Age at first birth, years (among parous women)** | | |  | |  | |
| <20 | 910 | (4.0) | 2,077 | (6.5) | 7,786 | (17.7) |
| 20-24 | 4,490 | (20.0) | 10,759 | (33.6) | 23,291 | (53.0) |
| 25-29 | 8,658 | (38.5) | 13,461 | (42.1) | 9,710 | (22.1) |
| ≥30 | 8,435 | (37.5) | 5,690 | (17.8) | 3,127 | (7.1) |
| Missing | 62 |  | 131 |  | 216 |  |
| **OC use** |  | |  | |  | |
| Never | 1,624 | (5.8) | 7,493 | (20.5) | 20,868 | (43.3) |
| Ever | 26,582 | (94.2) | 29,100 | (79.5) | 27,378 | (56.8) |
| Missing | 26 |  | 49 |  | 33 |  |
| **Current OC use (among women age <50)** | | |  | |  | |
| Never | 1,624 | (5.8) | N/A | | N/A | |
| Former | 22,897 | (81.2) |
| Current | 3,685 | (13.1) |
| Missing | 26 |  |
| **HRT use (among women age ≥50)** | | |  | |  | |
| Never | N/A | | 17,878 | (49.0) | 14,303 | (29.8) |
| Former | 6,191 | (34.0) | 7,459 | (15.5) |
| Current | 12,414 | (17.0) | 26,271 | (54.7) |
| Missing | 159 |  | 246 |  |
| **Type of HRT use (among current users age ≥50)** | | |  | |  | |
| Current E-type | N/A | | 2,976 | (52.7) | N/A | |
| Current C-type | 2,676 | (47.4) |
| Missing | 539 |  |
| **Age at menopause, years (among women age ≥50)** | | |  | |  | |
| <40 | N/A | | 419 | (1.9) | 6,718 | (14.0) |
| 40-44 | 1,482 | (6.8) | 6,728 | (14.0) |
| 45-49 | 5,171 | (23.8) | 11,156 | (23.3) |
| 50-54 | 11,563 | (53.2) | 17,688 | (36.9) |
| ≥55 | 3,094 | (14.2) | 5,624 | (11.7) |
| Missing | 14,913 |  | 365 |  |
| **Height, m** |  | |  | |  | |
| Median (range) | 1.7 | (1.2-2.0) | 1.6 | (1.3-2.1) | 1.6 | (1.2-2.0) |
| **Body mass index, kg/m2** |  | |  | |  | |
| <25 | 15,653 | (56.7) | 17,912 | (50.0) | 19,390 | (40.6) |
| ≥25 - <30 | 7,702 | (27.9) | 12,525 | (34.9) | 16,775 | (35.1) |
| ≥30 | 4,241 | (15.4) | 5,423 | (15.1) | 11,592 | (24.3) |
| Missing | 636 |  | 782 |  | 522 |  |
| **Alcohol, g/day** |  | |  | |  | |
| None | 4,227 | (15.0) | 7,359 | (20.5) | 17 | (0.1) |
| <5 | 3,718 | (13.2) | 5,000 | (13.9) | 17,767 | (74.4) |
| 5-14 | 8,374 | (29.7) | 10,173 | (28.4) | 2,900 | (12.2) |
| 15-24 | 5,775 | (20.5) | 7,032 | (19.6) | 1,697 | (7.1) |
| 25-34 | 2,948 | (10.5) | 3,282 | (9.1) | 405 | (1.7) |
| 35-44 | 1,526 | (5.4) | 1,533 | (4.3) | 628 | (2.6) |
| ≥45 | 1,621 | (5.8) | 1,496 | (4.2) | 460 | (1.9) |
| Missing | 43 |  | 767 |  | 24,405 |  |
| **History of BBD** |  | |  | |  | |
| No | 23,040 | (81.6) | 26,669 | (72.8) | 33,365 | (70.2) |
| Yes | 5,192 | (18.4) | 9,973 | (27.2) | 14,136 | (29.8) |
| Missing | 0 |  | 0 |  | 778 |  |
| **Breast cancer family history in first degree relatives** | | |  | |  | |
| No | 23,708 | (84.0) | 30,528 | (83.3) | 40,622 | (85.6) |
| Yes | 4,524 | (16.0) | 6,114 | (16.7) | 6,837 | (14.4) |
| Missing | 0 |  | 0 |  | 820 |  |
| \* Risk factor distributions are reported as n (%), unless otherwise specified. BMI and alcohol intake were assessed at baseline. BBD = benign breast disease, GS = Generations Study, C-type = estrogen and progestogen combined, E-type = estrogen-only, HRT = hormone replacement therapy, OC = oral contraceptive, PLCO = Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial. | | | | | | |

**Supplementary Table 2.** Parameters used for the development of the **iCARE-Lit** breast cancer risk prediction model\*

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **US Population distribution** | | | | **UK Population distribution** | | | | **Relative risk of breast cancer** | | | |
| **Breast cancer risk factors** | **Ages <50** | **Ages ≥50** | **Reference** | | **Ages <50** | **Ages ≥50** | | **Reference** | **Ages <50** | **Ages ≥50** | **Reference** | |
| **Age at menarche, years** |  |  |  | |  |  | |  |  |  |  | |
| ≤10 | 15.0% | 15.2% | 30 | | 12.2% | 11.7% | | 1 | 1.19 | 1.19 | 31 | |
| 11 | 18.5% | 18.1% | 21.6% | 21.5% | | 1.09 | 1.09 |
| 12 | 22.8% | 22.2% | 26.8% | 26.5% | | 1.07 | 1.07 |
| 13 | 21.4% | 21.6% | 21.5% | 21.3% | | Ref. | Ref. |
| 14 | 13.7% | 14.1% | 11.5% | 12.1% | | 0.98 | 0.98 |
| 15 | 6.1% | 6.1% | 4.4% | 4.8% | | 0.92 | 0.92 |
| ≥16 | 2.6% | 2.6% | 1.9% | 2.0% | | 0.82 | 0.82 |
| **Age at menopause, years** |  |  |  | |  |  | |  |  |  |  | |
| <40 | - | 21.3% | 32 | | - | 2.2% | | 1 | - | 0.67 | 31 | |
| 40-44 | - | 13.3% | - | 7.0% | | - | 0.73 |
| 45-49 | - | 22.7% | - | 23.4% | | - | 0.86 |
| 50-54 | - | 30.3% | - | 53.3% | | - | Ref. |
| ≥55 | - | 12.4% | - | 14.0% | | - | 1.12 |
| **Parity** |  |  |  | |  |  | |  |  |  |  | |
| Nulliparous | 20.4% | 17.6% | 30 | | 28.0% | 13.0% | | 33 | Ref. | Ref. | 34 | |
| 1 birth | 18.5% | 15.1% | 16.0% | 13.8% | | 0.87 | 0.87 |
| 2 births | 32.2% | 33.7% | 33.0% | 37.6% | | 0.81 | 0.81 |
| 3+ births | 28.9% | 33.6% | 23.0% | 35.6% | | 0.71 | 0.71 |
| **Age at first birth, years (among parous women)** | | | |  |  | |  | |  |  | |  |
| <20 | 23.1% | 26.8% | 30 | | 16.7% | 15.9% | | 33 | Ref. | Ref. | 34 | |
| 20-24 | 32.1% | 38.9% | 31.9% | 45.5% | | 1.01 | 1.01 |
| 25-29 | 25.3% | 21.7% | 32.0% | 27.3% | | 1.11 | 1.11 |
| ≥30 | 19.5% | 12.6% | 19.4% | 11.4% | | 1.24 | 1.24 |
| **OC use** |  |  |  | |  |  | |  |  |  |  | |
| Never | 14.3% | 19.1% | 35 | | 18.5% | 40.9% | | 36,37 | Ref. | Ref. | 38 | |
| Ever |  | 80.9% |  | 59.1% | | 1.14 | 1.14 |
| Former | 72.6% | - | 68.4% | - | | 1.12 | - |
| Current | 13.1% | - | 13.1% | - | | 1.33 | - |
| **BMI, kg/m2** |  |  |  | |  |  | |  |  |  |  | |
| <18.5 | 1.5% | - | 30 | | 3.3% | - | | 36,37 | 1.28 | - | 39 | |
| 18.5 - <25 | 41.0% | - | 35.7% | - | | Ref. | - |
| 25 - <30 | 29.9% | - | 35.8% | - | | 0.92 | - |
| ≥30 | 27.6% | - | 25.2% | - | | 0.74 | - |
| **Height, cm†** |  |  | 35 | |  |  | | 36,37 |  |  |  | |
| mean (sd) | 162.3 (6.4) | 162.3 (6.4) | 162.9 (6.5) | 159.6 (6.5) | | 1.17 | 1.17 | 40 | |
| **Alcohol, g/day** |  |  |  | |  |  | |  |  |  |  | |
| None | 41.8% | 45.0% | 32 | | 10.6% | 12.7% | | 41 | Ref. | Ref. | 42 | |
| <5 | 42.2% | 39.7% | 30.7% | 37.2% | | 1.01 | 1.01 |
| 5-14 | 12.4% | 12.1% | 27.6% | 23.8% | | 1.03 | 1.03 |
| 15-24 | 2.6% | 2.3% | 13.3% | 13.6% | | 1.13 | 1.13 |
| 25-34 | 0.6% | 0.6% | 8.0% | 5.3% | | 1.21 | 1.21 |
| 35-44 | 0.3% | 0.2% | 3.6% | 3.3% | | 1.32 | 1.32 |
| ≥45 | 0.2% | 0.2% | 6.1% | 4.1% | | 1.46 | 1.46 |
| **History of BBD** |  |  |  | |  |  | |  |  |  |  | |
| No | 83.3% | 83.2% | 43 | | 83.1% | 83.5% | | 43 | Ref. | Ref. | 44 | |
| Yes | 16.7% | 16.8% | 16.9% | 16.5% | | 1.68 | 1.51 |
| **Family history of breast cancer in first degree relatives** | | | |  |  | |  | |  |  | |  |
| No | 93.0% | 86.4% | 30 | | 93.1% | 90.5% | | 45 | Ref. | Ref. | 46 | |
| Yes | 7.0% | 13.6% | 6.9% | 9.5% | | 2.50 | 1.60 |
| I**nteraction between BMI and HRT use** | | | |  |  | |  | |  |  | |  |
| BMI <25: Never use | - | 20.4% | 30 | | - | 18.1% | | 1,36,37,47 | - | Ref | 48 | |
| BMI 25-29: Never use | - | 18.4% | - | 23.1% | | - | 1.13 |
| BMI ≥30: Never use | - | 19.5% | - | 19.4% | | - | 1.25 |
|  |  |  |  |  | |  |  |
| BMI <25: Former use | - | 10.5% | - | 9.2% | | - | 1.00 |
| BMI 25-29: Former use | - | 10.2% | - | 11.3% | | - | 1.13 |
| BMI ≥30: Former use | - | 11.1% | - | 10.1% | | - | 1.25 |
|  |  |  |  |  | |  |  |
| BMI <25: Current C-type | - | 1.7% | - | 2.3% | | - | 2.19 |
| BMI 25-29: Current C-type | - | 1.4% | - | 3.1% | | - | 2.18 |
| BMI ≥30: Current C-type | - | 1.1% | - | 2.2% | | - | 2.11 |
|  |  |  |  |  | |  |  |
| BMI <25: Current E-type | - | 2.5% | - | 0.4% | | - | 1.57 |
| BMI 25-29: Current E-type | - | 1.9% | - | 0.4% | | - | 1.49 |
| BMI ≥30: Current E-type | - | 1.3% | - | 0.3% | | - | 1.45 |
| \* BBD = benign breast disease, BMI = body mass index, C-type = estrogen and progestogen combined, E-type = estrogen-only, HRT = hormone replacement therapy, OC = oral contraceptive.  † Relative risk for is reported per 10 cm | | | | | | | | | | |  | |

**Supplementary Table 3.** Parameters used for the development of the **iCARE-BPC3** breast cancer risk prediction model\*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **US Population** | | | **UK Population** | | | | **Relative risk of breast cancer** |
| **Breast cancer risk factors** | | **Distribution** | **Reference** | | **Distribution** | | **Reference** | |
| **Age at menarche, years** | | | |  |  | | | |  |
| ≤11 | | 15.2% | 35 | | 11.7% | | 1 | | 1.09 |
| >11-11.5 | | 7.8% | 9.4% | | 1.14 |
| >11.5-12 | | 10.3% | 12.2% | | 1.05 |
| >12-13 | | 22.2% | 26.5% | | Ref. |
| >13-14 | | 21.6% | 21.3% | | 0.98 |
| >14-15 | | 14.1% | 12.1% | | 1.00 |
| >15 | | 8.7% | 6.8% | | 0.95 |
| **Age at menopause, years** | | | |  |  | | | |  |
| ≤40 | | 25.5% | 35 | | 0.7% | | 1 | | Ref. |
| >40-45 | | 15.0% | 10.1% | | 0.98 |
| >45-47 | | 6.0% | 10.6% | | 0.99 |
| >47-48 | | 5.6% | 7.3% | | 1.07 |
| >48-50 | | 16.9% | 16.4% | | 1.14 |
| >50-51 | | 4.4% | 9.6% | | 1.15 |
| >51-52 | | 5.9% | 9.1% | | 1.23 |
| >52-53 | | 4.9% | 7.6% | | 1.29 |
| >53-55 | | 9.3% | 13.4% | | 1.29 |
| >55 | | 6.3% | 15.3% | | 1.12 |
| **Parity** |  |  | | |  |  | | |  |
| Nulliparous | | 17.6% | 30 | | 13.0% | | 33 | | Ref. |
| 1 birth | | 15.1% | 13.8% | | 0.84 |
| 2 births | | 33.7% | 37.6% | | 0.77 |
| 3 births | | 19.6% | 20.8% | | 0.74 |
| 4+ births | | 14.0% | 14.8% | | 0.71 |
| **Age at first birth, years (among parous women)** | | | | |  | | | |  |
| ≤19 | | 26.8% | 30 | | 15.1% | | 33 | | Ref. |
| >19-22 | | 25.8% | 27.9% | | 1.06 |
| >22-23 | | 7.1% | 9.3% | | 1.03 |
| >23-25 | | 12.0% | 16.3% | | 1.02 |
| >25-27 | | 9.4% | 11.7% | | 1.21 |
| >27-30 | | 9.2% | 10.4% | | 1.39 |
| >30-34 | | 6.2% | 5.9% | | 1.37 |
| >34-38 | | 2.6% | 2.4% | | 1.49 |
| >38 | | 1.0% | 1.2% | | 1.28 |
| **Height, meters** | | | | |  | | | |  |
| ≤1.55 | | 12.9% | 35 | | 23.8% | | 36,37 | | Ref. |
| >1.55-1.57 | | 7.6% | 11.0% | | 1.11 |
| >1.57-1.60 | | 15.7% | 17.1% | | 1.13 |
| >1.60-1.61 | | 5.5% | 6.4% | | 1.15 |
| >1.61-1.63 | | 12.4% | 11.0% | | 1.08 |
| >1.63-1.65 | | 11.7% | 10.3% | | 1.20 |
| >1.65-1.66 | | 5.4% | 3.9% | | 1.18 |
| >1.66-1.68 | | 9.9% | 6.6% | | 1.28 |
| >1.68-1.71 | | 10.3% | 6.0% | | 1.19 |
| >1.71 | | 8.5% | 4.0% | | 1.28 |
| **Alcohol, drinks/week** | | | | |  | | | |  |
| None | | 44.9% | 32 | | 13.2% | | 41 | | Ref. |
| >0-0.4 | | 7.2% | 1.0% | | 0.94 |
| >0.4-0.8 | | 6.3% | 3.3% | | 1.01 |
| >0.8-1.5 | | 8.6% | 8.8% | | 1.04 |
| >1.5-3.2 | | 10.2% | 20.4% | | 0.99 |
| >3.2-5.7 | | 9.4% | 18.8% | | 1.05 |
| >5.7-9.8 | | 6.9% | 15.9% | | 1.15 |
| >9.8 | | 6.4% | 18.5% | | 1.22 |
| **Breast cancer family history in first degree relatives** | | | | |  | | | |  |
| No | | 86.4% | 30 | | 90.5% | | 45 | | Ref. |
| Yes | | 13.6% | 9.5% | | 1.43 |
| **Smoking status** | | | | |  | | | |  |
| Never | | 55.2% | 30 | | 48.3% | | 36,37 | | Ref. |
| Ever | | 44.8% | 51.7% | | 1.09 |
| **BMI, kg/m2** | |  |  | |  | |  | |  |
| ≤21.5 | | 11.0% | 35 | | 9.1% | |  | | Ref. |
| >21.5-23 | | 9.3% | 7.5% | |  | | 1.04 |
| >23-24.2 | | 8.1% | 7.7% | |  | | 0.89 |
| >24.2-25.3 | | 8.8% | 7.9% | |  | | 0.95 |
| >25.3-26.5 | | 7.9% | 9.6% | | 1,36,37 | | 0.85 |
| >26.5-27.8 | | 9.2% | 10.3% | |  | | 1.06 |
| >27.8-29.3 | | 9.8% | 10.6% | |  | | 1.03 |
| >29.3-31.4 | | 9.5% | 12.8% | |  | | 0.92 |
| >31.4-34.6 | | 11.5% | 13.0% | |  | | 0.94 |
| >34.6 | | 14.8% | 11.4% | |  | | 1.06 |
| **HRT use, C-type** | | | | |  | | | |  |
| Otherwise | | 90.2% | 35 | | 75.7% | | 49 | | Ref. |
| Postmenopausal and ever use | | 9.8% | 24.3% | | 1.29 |
| **HRT use, E-type** | | | | |  | | |  |  |
| Otherwise | | 74.4% | 35 | | 86.9% | | 49 | | Ref. |
| Postmenopausal and ever use | | 25.6% | 13.1% | | 1.02 |
| **Current HRT use** | | | | |  | | | |  |
| Otherwise | | 90.2% | 35 | | 91.3% | | 36,37 | | Ref. |
| Postmenopausal and current use | | 9.8% | 8.7% | | 1.22 |
| I**nteraction between BMI and HRT use** | | | | |  | | | |  |
| BMI ≤21.5 or Premenopausal | | 17.9% | 30,35 | | 13.8% | | 1,36,37 | | Ref. |
| BMI >21.5-23: Never use | | 5.4% | 4.3% | | 1.02 |
| BMI >23-24.2: Never use | | 4.7% | 4.5% | | 1.20 |
| BMI >24.2-25.3: Never use | | 5.4% | 4.6% | | 1.24 |
| BMI >25.3-26.5: Never use | | 4.7% | 5.5% | | 1.40 |
| BMI >26.5-27.8: Never use | | 5.5% | 5.8% | | 1.07 |
| BMI >27.8-29.3: Never use | | 5.8% | 6.2% | | 1.28 |
| BMI >29.3-31.4: Never use | | 5.7% | 7.2% | | 1.53 |
| BMI >31.4-34.6: Never use | | 6.5% | 7.6% | | 1.58 |
| BMI >34.6: Never use | | 8.9% | 6.6% | | 1.45 |
| BMI >21.5-23: Ever use | | 2.9% | 2.7% | | 1.01 |
| BMI >23-24.2: Ever use | | 2.6% | 2.9% | | 1.16 |
| BMI >24.2-25.3: Ever use | | 2.7% | 3.1% | | 1.19 |
| BMI >25.3-26.5: Ever use | | 2.6% | 3.6% | | 1.27 |
| BMI >26.5-27.8: Ever use | | 3.2% | 4.0% | | 1.11 |
| BMI >27.8-29.3: Ever use | | 3.1% | 3.6% | | 1.12 |
| BMI >29.3-31.4: Ever use | | 3.1% | 5.0% | | 1.40 |
| BMI >31.4-34.6: Ever use | | 4.1% | 4.8% | | 1.27 |
| BMI >34.6: Ever use | | 4.9% | 4.3% | | 1.29 |
| \* BMI = body mass index, C-type = estrogen and progestogen combined, E-type = estrogen-only, HRT = hormone replacement therapy. | | | | | | | | | |

**Supplementary Table 4.** Risk factors included in the breast cancer risk prediction models\*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Breast cancer risk models** | | | |
| **Risk factors included in the model** | **iCARE-Lit** | **iCARE-BPC3** | **BCRAT** | **IBIS** |
| Age at menarche | ✔ | ✔ | ✔ | ✔ |
| Parity | ✔ | ✔ |  | ✔ |
| Age at first birth | ✔ | ✔ | ✔ | ✔ |
| Oral contraceptive use | ✔ |  |  |  |
| Body mass index (BMI) | ✔ | ✔ |  | ✔ |
| Alcohol use | ✔ | ✔ |  |  |
| Smoking status |  | ✔ |  |  |
| Menopausal status† |  | ✔ |  | ✔ |
| Age at menopause | ✔ | ✔ |  | ✔ |
| Menopausal replacement therapy use | ✔ | ✔ |  | ✔ |
| BBD‡ | ✔ |  | ✔ |  |
| Type of BBD |  |  |  |  |
| BBD - hyperplasia |  |  |  | ✔ |
| BBD - atypical hyperplasia |  |  | ✔ | ✔ |
| Lobular Carcinoma *in situ* (LCIS) |  |  |  | ✔ |
| Family history of breast and/or ovarian cancer |  |  |  |  |
| First-degree relatives with breast cancer (yes/no) | ✔ | ✔ | ✔ | ✔ |
| Number of first-degree relatives with breast cancer |  |  | ✔ | ✔ |
| Second-degree relatives with breast cancer |  |  |  | ✔ |
| Third-degree relatives with breast cancer |  |  |  | ✔ |
| Age of onset of breast cancer in a relative |  |  |  | ✔ |
| Bilateral breast cancer in a relative |  |  |  | ✔ |
| Ovarian cancer in a relative |  |  |  | ✔ |
| Male breast cancer |  |  |  | ✔ |
| \* BBD = benign breast disease.  † iCARE-Lit and BCRAT use age 50 as a surrogate for menopausal status ‡ Assessed as yes/no for iCARE-Lit and as number of biopsies for BCRAT | | | | |

**Supplementary Table 5.** 313 Single nucleotide polymorphisms included in the polygenic risk score\*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Chromosome** | **Position** | **RS ID** | **Reference Allele** | **Effect Allele** | **EAF** | **Log OR** | |
| 1 | 100880328 | rs612683 | A | T | 0.4097 | 0.0373 | |
| 1 | 10566215 | rs616488 | A | G | 0.329 | -0.0586 | |
| 1 | 110198129 | 1:110198129\_CAAA\_C | CAAA | C | 0.7755 | 0.0458 | |
| 1 | 114445880 | rs7513707 | G | A | 0.1664 | 0.0621 | |
| 1 | 118141492 | rs12406858 | A | C | 0.2657 | 0.0452 | |
| 1 | 120257110 | rs637868 | T | C | 0.5309 | 0.0385 | |
| 1 | 121280613 | rs11249433 | A | G | 0.4053 | 0.0881 | |
| 1 | 121287994 | rs111458676 | A | G | 0.106 | -0.0673 | |
| 1 | 145604302 | rs143384623 | C | CT | 0.3515 | -0.0399 | |
| 1 | 149906413 | rs11205303 | T | C | 0.4017 | 0.0548 | |
| 1 | 155556971 | rs12091730 | G | A | 0.2302 | 0.0499 | |
| 1 | 168171052 | rs761575824 | CA | C | 0.1097 | -0.068 | |
| 1 | 172328767 | rs11463354 | T | TA | 0.3305 | -0.0435 | |
| 1 | 18807339 | rs2992756 | T | C | 0.5145 | -0.0564 | |
| 1 | 201437832 | rs35383942 | C | T | 0.0559 | 0.0917 | |
| 1 | 202184600 | rs6686987 | C | T | 0.3992 | -0.0065 | |
| 1 | 203770448 | rs7514172 | T | A | 0.2715 | 0.0498 | |
| 1 | 204502514 | rs11268668 | T | TTCTGAAACAGGG | 0.8028 | -0.0321 | |
| 1 | 208076291 | rs2785646 | G | A | 0.3337 | -0.0366 | |
| 1 | 217053815 | rs2576261 | T | G | 0.328 | 0.0417 | |
| 1 | 217220574 | rs11117758 | G | A | 0.2107 | -0.044 | |
| 1 | 220671050 | rs11118563 | C | T | 0.2415 | 0.0418 | |
| 1 | 242034263 | rs72755295 | A | G | 0.0305 | 0.1428 | |
| 1 | 41380440 | rs4233486 | C | T | 0.6438 | 0.0426 | |
| 1 | 41389220 | rs114282204 | T | C | 0.0169 | 0.155 | |
| 1 | 46670206 | 1:46670206\_TC\_T | TC | T | 0.2973 | 0.0447 | |
| 1 | 51467096 | 1:51467096\_CT\_C | CT | C | 0.48 | 0.0374 | |
| 1 | 7917076 | rs707475 | G | A | 0.3899 | -0.0409 | |
| 1 | 88156923 | rs17426269 | G | A | 0.1487 | 0.0494 | |
| 1 | 88428199 | rs2151842 | C | A | 0.2477 | -0.0387 | |
| 2 | 10138983 | rs78425380 | T | C | 0.116 | 0.0603 | |
| 2 | 121058254 | rs6746250 | A | G | 0.7047 | -0.0334 | |
| 2 | 121089731 | rs17625845 | T | C | 0.1943 | -0.0427 | |
| 2 | 121159205 | rs10164550 | G | A | 0.3527 | -0.044 | |
| 2 | 121246568 | rs10179592 | T | C | 0.897 | 0.0992 | |
| 2 | 172974566 | rs17726078 | C | G | 0.4743 | -0.0473 | |
| 2 | 174212910 | rs1550622 | A | G | 0.845 | 0.0593 | |
| 2 | 192381934 | rs2356656 | C | T | 0.8588 | 0.0316 | |
| 2 | 19315675 | rs6743383 | T | A | 0.5599 | -0.0331 | |
| 2 | 202204741 | rs10197246 | T | C | 0.721 | -0.0492 | |
| 2 | 217920769 | rs4442975 | G | T | 0.5001 | -0.1318 | |
| 2 | 217955896 | 2:217955896\_GA\_G | GA | G | 0.0382 | -0.2016 | |
| 2 | 218292158 | rs11693806 | C | G | 0.7309 | -0.0757 | |
| 2 | 218714845 | rs3791977 | G | A | 0.3917 | -0.0431 | |
| 2 | 241388857 | rs4676356 | C | A | 0.9772 | -0.1232 | |
| 2 | 25129473 | rs6725517 | A | G | 0.4082 | -0.0427 | |
| 2 | 29179452 | rs12472404 | G | C | 0.2287 | -0.0066 | |
| 2 | 29615233 | rs4322799 | T | C | 0.2622 | -0.0427 | |
| 2 | 39699510 | rs553796823 | C | CT | 0.4659 | -0.0402 | |
| 2 | 70172587 | rs6756513 | G | A | 0.2787 | -0.0412 | |
| 2 | 88358825 | rs1036759 | G | C | 0.3081 | 0.0473 | |
| 3 | 141112859 | 3:141112859\_CTT\_C | CTT | C | 0.4149 | 0.0551 | |
| 3 | 172285237 | rs58058861 | G | A | 0.2131 | 0.0422 | |
| 3 | 189774456 | rs9882792 | C | T | 0.2235 | -0.0478 | |
| 3 | 27353716 | rs552647 | C | A | 0.5259 | 0.0748 | |
| 3 | 27388664 | rs62255657 | C | G | 0.2735 | 0.0502 | |
| 3 | 29294845 | rs112476261 | C | T | 0.0163 | -0.1281 | |
| 3 | 30684907 | rs17838698 | C | T | 0.2975 | 0.0592 | |
| 3 | 46888198 | rs56387622 | T | C | 0.1032 | -0.0806 | |
| 3 | 4742251 | rs6762558 | A | G | 0.3802 | 0.0616 | |
| 3 | 49709912 | rs371314787 | C | CT | 0.2873 | -0.0367 | |
| 3 | 55970777 | rs138866686 | A | AT | 0.0305 | -0.1195 | |
| 3 | 59373745 | rs2886671 | C | T | 0.4294 | -0.0394 | |
| 3 | 63887449 |  | T | TTG | 0.1297 | 0.0648 | |
| 3 | 71620370 | rs9825432 | T | G | 0.6382 | -0.0374 | |
| 3 | 87037543 | rs13066793 | A | G | 0.0921 | -0.0723 | |
| 3 | 99403877 | rs639355 | G | A | 0.4852 | -0.0376 | |
| 4 | 106069013 | rs62331150 | G | T | 0.2289 | 0.0471 | |
| 4 | 126752992 |  | A | AAT | 0.5167 | -0.0377 | |
| 4 | 143467195 | rs56039025 | C | T | 0.1115 | -0.0569 | |
| 4 | 151218296 | rs745707748 | CATATTT | C | 0.6533 | 0.0388 | |
| 4 | 175842495 | rs28436676 | G | A | 0.1161 | -0.0898 | |
| 4 | 175847436 | rs62334414 | C | A | 0.3433 | 0.0348 | |
| 4 | 187503758 |  | A | T | 0.4471 | 0.0357 | |
| 4 | 38784633 | rs10012017 | G | T | 0.2493 | 0.0489 | |
| 4 | 84370124 | rs774021038 | TAA | TA | 0.5324 | -0.0464 | |
| 4 | 89240476 | rs17014016 | G | A | 0.4395 | 0.0352 | |
| 4 | 92594859 | rs775780062 | TTCTTTC | T | 0.4445 | -0.0407 | |
| 5 | 104300273 | rs17157372 | G | T | 0.181 | -0.0487 | |
| 5 | 122478676 | rs335160 | C | A | 0.7448 | -0.0386 | |
| 5 | 122705244 | rs1428387 | C | T | 0.0306 | 0.0944 | |
| 5 | 1279790 | rs10069690 | C | T | 0.2592 | 0.0617 | |
| 5 | 1296255 | rs3215401 | A | AG | 0.3072 | -0.0549 | |
| 5 | 131640536 | rs6860806 | A | G | 0.5427 | 0.0392 | |
| 5 | 132407058 | rs6596100 | C | T | 0.245 | -0.0388 | |
| 5 | 1353077 | rs62329727 | T | C | 0.0121 | 0.1552 | |
| 5 | 158244083 | rs1432679 | C | T | 0.5683 | -0.0677 | |
| 5 | 16231194 | rs17611291 | G | C | 0.5594 | -0.0426 | |
| 5 | 169591460 | rs10074269 | T | C | 0.3345 | 0.0412 | |
| 5 | 173358154 | rs6864691 | G | A | 0.4074 | 0.0365 | |
| 5 | 176134882 | rs4868701 | T | C | 0.5422 | 0.0363 | |
| 5 | 2777029 | rs4866496 | G | A | 0.4139 | 0.0391 | |
| 5 | 32579616 | rs770436441 | TCA | T | 0.4844 | 0.0363 | |
| 5 | 345109 | rs62641919 | T | C | 0.0544 | 0.084 | |
| 5 | 44508264 | rs138335056 | G | GT | 0.1265 | -0.1177 | |
| 5 | 44619502 | rs187108781 | A | G | 0.1549 | -0.1101 | |
| 5 | 44649944 | rs4613718 | C | T | 0.601 | 0.0492 | |
| 5 | 44706498 | rs10941679 | A | G | 0.2481 | 0.0497 | |
| 5 | 44853593 | rs17343002 | G | C | 0.3081 | -0.0336 | |
| 5 | 52679539 |  | C | CA | 0.0998 | 0.0571 | |
| 5 | 55662540 | rs553874618 | C | CT | 0.3631 | -0.0458 | |
| 5 | 55965167 | rs889310 | C | T | 0.5576 | 0.0394 | |
| 5 | 56023083 | rs16886165 | T | G | 0.1583 | 0.1366 | |
| 5 | 56042972 | rs76250845 | C | T | 0.0521 | 0.0865 | |
| 5 | 56045081 | rs11949391 | T | C | 0.1655 | -0.0564 | |
| 5 | 58241712 | rs113778879 | C | T | 0.575 | -0.0434 | |
| 5 | 71965007 | rs3010266 | G | A | 0.2572 | -0.041 | |
| 5 | 73234583 | rs157557 | T | C | 0.3213 | -0.0363 | |
| 5 | 77155397 | rs767431357 | GT | G | 0.3466 | -0.0408 | |
| 5 | 79180995 | rs34525310 | G | GA | 0.1755 | 0.0328 | |
| 5 | 81512947 | 5:81512947\_TA\_T | TA | T | 0.2503 | -0.0598 | |
| 5 | 90789470 | rs332529 | G | A | 0.158 | -0.0564 | |
| 6 | 130341728 | rs55941023 | C | CT | 0.7116 | 0.0472 | |
| 6 | 13713366 | rs418053 | G | C | 0.5691 | -0.0553 | |
| 6 | 149595505 | rs2121348 | T | C | 0.2061 | -0.0476 | |
| 6 | 151949806 | rs6913578 | A | C | 0.3083 | 0.0703 | |
| 6 | 151955914 | rs60954078 | A | G | 0.0713 | 0.1449 | |
| 6 | 152022664 | 6:152022664\_CAAAAAAA\_C | CAAAAAAA | C | 0.6119 | 0.0137 | |
| 6 | 152023191 | rs851984 | G | A | 0.3965 | 0.0626 | |
| 6 | 152055978 | rs6904031 | A | T | 0.0627 | 0.074 | |
| 6 | 152432902 | rs910416 | C | T | 0.5146 | 0.0649 | |
| 6 | 16399557 | rs3819405 | C | T | 0.3299 | -0.0373 | |
| 6 | 169006947 | rs9364472 | C | G | 0.5202 | -0.0308 | |
| 6 | 170332621 | rs6940159 | T | C | 0.6158 | 0.0373 | |
| 6 | 18783140 | rs12211970 | G | A | 0.62 | 0.0326 | |
| 6 | 20537845 | rs769485514 | CA | C | 0.4733 | -0.0391 | |
| 6 | 21923810 | rs9358466 | T | C | 0.4303 | -0.0321 | |
| 6 | 27425644 | rs34196306 | G | C | 0.0815 | -0.0737 | |
| 6 | 43227141 | rs111342015 | G | A | 0.0985 | -0.064 | |
| 6 | 82263549 | rs10623112 | AAT | A | 0.4262 | 0.0477 | |
| 6 | 85912194 | 6:85912194\_CAA\_C | CAA | C | 0.0604 | 0.0762 | |
| 6 | 87803819 | rs73754909 | T | C | 0.277 | 0.0383 | |
| 7 | 101552440 | rs71559437 | G | A | 0.1255 | -0.0568 | |
| 7 | 102481842 | rs7800548 | T | C | 0.3416 | 0.0418 | |
| 7 | 130656911 | rs12706954 | C | T | 0.3734 | -0.0476 | |
| 7 | 130674481 | rs68056147 | G | A | 0.2971 | 0.0416 | |
| 7 | 139943702 | rs201664599 | CT | C | 0.5381 | 0.0582 | |
| 7 | 144048902 | rs62485509 | G | T | 0.2284 | -0.0563 | |
| 7 | 21940960 | rs7971 | A | G | 0.3515 | -0.0467 | |
| 7 | 25569548 | rs289997 | C | T | 0.1667 | -0.0486 | |
| 7 | 28869017 | rs74765302 | G | A | 0.1072 | -0.0572 | |
| 7 | 55192256 | rs13244925 | A | C | 0.5497 | -0.0349 | |
| 7 | 91459189 |  | A | ATT | 0.3286 | 0.0452 | |
| 7 | 94113799 | rs17268829 | T | C | 0.2792 | 0.0449 | |
| 7 | 98005235 | rs4439053 | G | A | 0.1627 | -0.0467 | |
| 7 | 99948655 | rs111963714 | T | G | 0.2109 | 0.042 | |
| 8 | 102483100 | rs62517052 | T | C | 0.0967 | 0.0593 | |
| 8 | 106358620 | rs12546444 | A | T | 0.1003 | -0.0745 | |
| 8 | 117209548 | rs13267382 | A | G | 0.6445 | -0.0417 | |
| 8 | 120862186 | rs62526620 | A | G | 0.1318 | 0.0527 | |
| 8 | 124563705 | rs35542655 | T | C | 0.1458 | 0.0477 | |
| 8 | 124571581 | rs12541094 | G | A | 0.4173 | 0.034 | |
| 8 | 124739913 | rs7842619 | T | G | 0.3985 | 0.0466 | |
| 8 | 128213561 | rs35961416 | C | CA | 0.4153 | -0.043 | |
| 8 | 128370949 | rs12550713 | C | G | 0.402 | 0.0642 | |
| 8 | 128372172 | rs10096351 | A | G | 0.5446 | 0.0597 | |
| 8 | 129199566 | rs1016578 | G | A | 0.1717 | 0.0615 | |
| 8 | 143669254 | rs7830152 | A | G | 0.339 | -0.0346 | |
| 8 | 170692 | rs66823261 | T | C | 0.2227 | 0.0477 | |
| 8 | 17787610 | 8:17787610\_CT\_C | CT | C | 0.623 | -0.0377 | |
| 8 | 23447496 | rs1028016 | A | G | 0.6487 | -0.0389 | |
| 8 | 23663653 | rs310295 | C | A | 0.4032 | 0.0335 | |
| 8 | 29509616 | rs9693444 | A | C | 0.6756 | -0.0601 | |
| 8 | 36858483 | rs13365225 | A | G | 0.182 | -0.076 | |
| 8 | 76230943 | rs1511243 | A | G | 0.8282 | 0.0755 | |
| 8 | 76333056 | rs72658084 | C | T | 0.0878 | 0.1129 | |
| 8 | 76378165 | rs1533366 | G | T | 0.3595 | -0.0391 | |
| 9 | 110303808 | 9:110303808\_TAA\_T | TAA | T | 0.2065 | 0.0797 | |
| 9 | 110837073 | rs10816625 | A | G | 0.063 | 0.1158 | |
| 9 | 110837176 | rs13294895 | C | T | 0.175 | 0.0653 | |
| 9 | 110849525 | rs7848334 | G | T | 0.5977 | 0.0153 | |
| 9 | 110885479 | rs630965 | C | T | 0.6222 | 0.0877 | |
| 9 | 119313486 | rs1895062 | A | G | 0.4087 | -0.0462 | |
| 9 | 129424719 | rs3861871 | A | G | 0.4577 | -0.0382 | |
| 9 | 136146597 | rs550057 | C | T | 0.2727 | 0.04 | |
| 9 | 21964882 | rs745322029 | CAAAA | C | 0.3184 | 0.055 | |
| 9 | 22041998 | rs17694493 | C | G | 0.1393 | 0.0289 | |
| 9 | 36928288 | rs4880038 | T | C | 0.5349 | 0.0249 | |
| 9 | 6880263 | rs10975870 | A | G | 0.2855 | 0.0348 | |
| 9 | 87782211 | rs665889 | T | C | 0.5094 | 0.0361 | |
| 9 | 98362587 | rs10120432 | T | C | 0.094 | 0.0576 | |
| 10 | 114777670 | rs10885405 | C | T | 0.4631 | 0.0472 | |
| 10 | 115128491 | rs12250948 | T | C | 0.7846 | -0.0592 | |
| 10 | 123095209 | rs9421410 | G | A | 0.3269 | -0.0538 | |
| 10 | 123340107 | rs45631580 | A | G | 0.0656 | 0.1508 | |
| 10 | 123340431 | 10:123340431\_GC\_G | GC | G | 0.5963 | -0.2408 | |
| 10 | 123349324 | rs45631563 | A | T | 0.0484 | -0.2609 | |
| 10 | 13892298 | rs10796139 | G | A | 0.4376 | 0.0371 | |
| 10 | 22032942 | rs7072776 | A | G | 0.7085 | -0.058 | |
| 10 | 22477776 | rs762131501 | ACC | A | 0.0202 | 0.1687 | |
| 10 | 22861490 | rs10764337 | A | C | 0.937 | 0.0875 | |
| 10 | 38523626 | rs2384736 | C | A | 0.3698 | 0.0404 | |
| 10 | 5794652 | rs55910451 | A | G | 0.2137 | 0.047 | |
| 10 | 64299890 | rs10995201 | A | G | 0.1603 | -0.1345 | |
| 10 | 64819996 | rs6479868 | G | T | 0.1958 | 0.0472 | |
| 10 | 71335574 | rs111833376 | C | T | 0.3179 | -0.0404 | |
| 10 | 80851257 | rs719338 | G | T | 0.6172 | -0.0805 | |
| 10 | 80886726 | rs4980029 | A | G | 0.1631 | 0.0762 | |
| 10 | 95292187 | 10:95292187\_CAA\_C | CAA | C | 0.8234 | -0.0512 | |
| 11 | 103614438 | rs7125780 | T | G | 0.6572 | 0.0147 | |
| 11 | 108267402 | rs199504893 | C | CA | 0.4173 | -0.0022 | |
| 11 | 111696440 | rs610437 | T | C | 0.6221 | -0.0396 | |
| 11 | 116727936 | rs625145 | A | T | 0.2046 | -0.0423 | |
| 11 | 122966626 | rs7121616 | A | G | 0.2922 | -0.0383 | |
| 11 | 129243417 | rs7939702 | T | G | 0.862 | -0.0543 | |
| 11 | 129461016 | rs11822830 | A | G | 0.6016 | 0.0453 | |
| 11 | 18664241 | rs10832963 | T | G | 0.7293 | 0.0461 | |
| 11 | 1895708 | rs4980386 | C | A | 0.3924 | -0.0762 | |
| 11 | 42844441 | rs4472923 | C | T | 0.3279 | -0.0336 | |
| 11 | 433617 | rs7394715 | T | C | 0.7969 | -0.0437 | |
| 11 | 44368892 | rs10838267 | G | A | 0.5495 | 0.0374 | |
| 11 | 46318032 | rs77047825 | C | G | 0.0659 | -0.0748 | |
| 11 | 65553492 | rs12287832 | C | A | 0.1867 | 0.0425 | |
| 11 | 65572431 | rs10896047 | G | A | 0.4886 | -0.0347 | |
| 11 | 69328130 | rs35039974 | A | T | 0.213 | -0.0423 | |
| 11 | 69330983 | rs661204 | G | A | 0.125 | 0.1022 | |
| 11 | 69331418 | rs78540526 | C | T | 0.0753 | 0.1782 | |
| 11 | 803017 | rs6597981 | A | G | 0.5167 | 0.0457 | |
| 12 | 103097887 | rs7132703 | C | T | 0.1175 | 0.0546 | |
| 12 | 111600134 | rs11065822 | G | T | 0.3715 | -0.0442 | |
| 12 | 115108136 | rs1061657 | T | C | 0.2615 | 0.0465 | |
| 12 | 115796577 | rs11067551 | A | G | 0.1959 | -0.0428 | |
| 12 | 115835836 | rs2454399 | T | C | 0.4171 | -0.0813 | |
| 12 | 120832146 | rs206966 | C | T | 0.1593 | 0.0516 | |
| 12 | 14413931 | rs12422552 | G | C | 0.2619 | 0.0484 | |
| 12 | 28149568 | rs788458 | C | T | 0.117 | -0.062 | |
| 12 | 28174817 | rs7297051 | C | T | 0.2421 | -0.0856 | |
| 12 | 28347382 | rs11049431 | C | T | 0.2153 | -0.0521 | |
| 12 | 29140260 | rs1027113 | G | A | 0.9126 | 0.0647 | |
| 12 | 293626 | rs797736 | A | G | 0.3711 | 0.0401 | |
| 12 | 57146069 | rs2277339 | T | G | 0.1037 | -0.0579 | |
| 12 | 70798355 | rs2870876 | A | T | 0.181 | 0.0469 | |
| 12 | 83064195 | rs111622698 | G | GA | 0.0992 | 0.0671 | |
| 12 | 85004551 | rs10862899 | C | T | 0.4955 | 0.0348 | |
| 12 | 96027759 | rs17356907 | A | G | 0.2963 | -0.0867 | |
| 13 | 32839990 | rs56404467 | G | A | 0.0174 | 0.0424 | |
| 13 | 32972626 | rs11571833 | A | T | 0.0079 | 0.2687 | |
| 13 | 43501356 | rs9315973 | A | G | 0.8303 | 0.0517 | |
| 13 | 73806982 | rs12870942 | T | C | 0.3153 | 0.0345 | |
| 13 | 73960952 | rs2181965 | A | G | 0.7618 | 0.0399 | |
| 14 | 105213978 | rs4983544 | T | G | 0.4588 | 0.0399 | |
| 14 | 37128564 | rs34914085 | C | A | 0.2122 | -0.0733 | |
| 14 | 37228504 | rs2253012 | C | T | 0.4434 | 0.039 | |
| 14 | 68660428 | rs2588809 | T | C | 0.8345 | -0.0474 | |
| 14 | 68979835 | rs11624333 | T | C | 0.2581 | -0.0911 | |
| 14 | 91751788 | 14:91751788\_TC\_T | TC | T | 0.6934 | 0.038 | |
| 14 | 91841069 | rs941764 | A | G | 0.3444 | 0.0513 | |
| 14 | 93070286 | rs78440108 | C | T | 0.1709 | -0.0577 | |
| 15 | 100905819 | rs144767203 | A | C | 0.11 | -0.0608 | |
| 15 | 46680811 | rs187010898 | C | A | 0.0115 | -0.1973 | |
| 15 | 50694306 | rs4774565 | A | G | 0.3446 | -0.0417 | |
| 15 | 66630569 | rs8042593 | G | A | 0.6413 | -0.0369 | |
| 15 | 67457698 | rs35874463 | A | G | 0.0496 | 0.0782 | |
| 15 | 75750383 | rs8035987 | T | C | 0.2604 | -0.0413 | |
| 15 | 91512267 | rs2290202 | G | T | 0.1353 | -0.0589 | |
| 16 | 10706580 | rs34872983 | G | A | 0.0695 | -0.074 | |
| 16 | 23007047 | rs75753503 | G | T | 0.0236 | 0.1218 | |
| 16 | 4008542 | 16:4008542\_CAAAAA\_C | CAAAAA | C | 0.8213 | -0.0329 | |
| 16 | 4106788 | rs11076805 | C | A | 0.2643 | -0.03 | |
| 16 | 52538825 | rs35668161 | C | A | 0.2562 | 0.1147 | |
| 16 | 52599188 | rs4784227 | C | T | 0.2406 | 0.107 | |
| 16 | 53809123 | rs55872725 | C | T | 0.4201 | -0.0704 | |
| 16 | 53861139 | rs6499648 | C | T | 0.7604 | -0.0338 | |
| 16 | 53861592 | rs7184573 | G | A | 0.3663 | -0.0337 | |
| 16 | 54682064 | rs28539243 | G | A | 0.485 | 0.0477 | |
| 16 | 6963972 | rs12709163 | C | G | 0.7835 | 0.0354 | |
| 16 | 80648296 | rs7500067 | A | G | 0.2303 | 0.0839 | |
| 16 | 85145977 | rs9931038 | T | C | 0.4856 | -0.0211 | |
| 16 | 87086492 | rs12449271 | T | C | 0.2586 | -0.0469 | |
| 17 | 29168077 | rs79461387 | G | T | 0.2613 | -0.0568 | |
| 17 | 39251123 | rs150537328 | T | C | 0.0682 | 0.0799 | |
| 17 | 40127060 | rs11296 | T | C | 0.057 | 0.0174 | |
| 17 | 40485239 | rs17881320 | G | T | 0.0874 | -0.0571 | |
| 17 | 40744470 | rs149370081 | G | A | 0.0124 | 0.2017 | |
| 17 | 43212339 | rs545502941 | C | CT | 0.2284 | 0.0438 | |
| 17 | 44283858 | rs2668667 | G | A | 0.1895 | -0.054 | |
| 17 | 53209774 | rs2787486 | A | C | 0.3023 | -0.0793 | |
| 17 | 77781725 | rs745570 | A | G | 0.5038 | -0.0401 | |
| 18 | 11696613 | rs16976596 | C | T | 0.1379 | -0.0381 | |
| 18 | 20634253 | rs11665269 | C | T | 0.6403 | -0.0415 | |
| 18 | 24125857 | rs1111207 | T | C | 0.4214 | 0.0346 | |
| 18 | 24337424 | rs527616 | C | G | 0.6205 | 0.0455 | |
| 18 | 24518050 | 18:24518050\_AT\_A | AT | A | 0.2773 | -0.0599 | |
| 18 | 25407513 | rs8092192 | C | G | 0.7126 | 0.0399 | |
| 18 | 29981526 | rs72931898 | G | A | 0.0474 | -0.1058 | |
| 18 | 42411803 | rs9954058 | G | C | 0.0717 | -0.0877 | |
| 18 | 42888797 | rs9952980 | T | C | 0.3519 | -0.0542 | |
| 19 | 13249921 | rs117922601 | G | T | 0.0513 | 0.0956 | |
| 19 | 17393925 | rs56069439 | C | A | 0.2958 | 0.0378 | |
| 19 | 18569492 | rs10164323 | C | T | 0.3481 | -0.0719 | |
| 19 | 19517054 | rs140702307 | C | CGGGCG | 0.3537 | 0.0437 | |
| 19 | 44283031 | rs56681946 | T | C | 0.3519 | 0.0619 | |
| 19 | 46166073 | rs4399645 | T | C | 0.6074 | -0.036 | |
| 19 | 55816678 | rs1172821 | C | T | 0.3626 | -0.0359 | |
| 20 | 11379842 | rs1154723 | T | C | 0.9483 | 0.0844 | |
| 20 | 41613706 | rs6030585 | C | G | 0.7928 | 0.0315 | |
| 20 | 52296849 | rs13039563 | G | A | 0.24 | 0.044 | |
| 20 | 5948227 | rs16991615 | G | A | 0.0628 | 0.076 | |
| 21 | 16364756 | rs2822999 | T | G | 0.1732 | 0.0646 | |
| 21 | 16566350 | rs2823130 | A | G | 0.0873 | 0.0595 | |
| 21 | 16574455 | rs2403907 | C | A | 0.3167 | -0.0707 | |
| 21 | 47762932 | rs4818836 | G | A | 0.0355 | 0.0946 | |
| 22 | 19766137 | rs9798754 | C | T | 0.3798 | -0.0367 | |
| 22 | 29121087 | rs17879961 | A | G | 0.0054 | 0.1839 | |
| 22 | 29135543 | rs5997390 | G | A | 0.087 | 0.0654 | |
| 22 | 29203724 | rs34134147 | C | T | 0.0209 | 0.1405 | |
| 22 | 29551872 | rs132289 | A | G | 0.9846 | -0.1716 | |
| 22 | 38583315 |  | AAAAG | AAAAGAAAG | 0.2805 | -0.0471 | |
| 22 | 39343916 | rs5750715 | T | A | 0.2541 | 0.0407 | |
| 22 | 40904707 | 22:40904707\_CT\_C | CT | C | 0.1099 | 0.1148 | |
| 22 | 43433100 | rs9611990 | C | T | 0.1144 | -0.06 | |
| 22 | 45319953 | rs112855987 | G | A | 0.4166 | -0.0134 | |
| 22 | 46283297 | rs28512361 | G | A | 0.1117 | 0.0736 | |
| \* EAF = effect allele frequency, OR = odds ratio. | | | | | | |

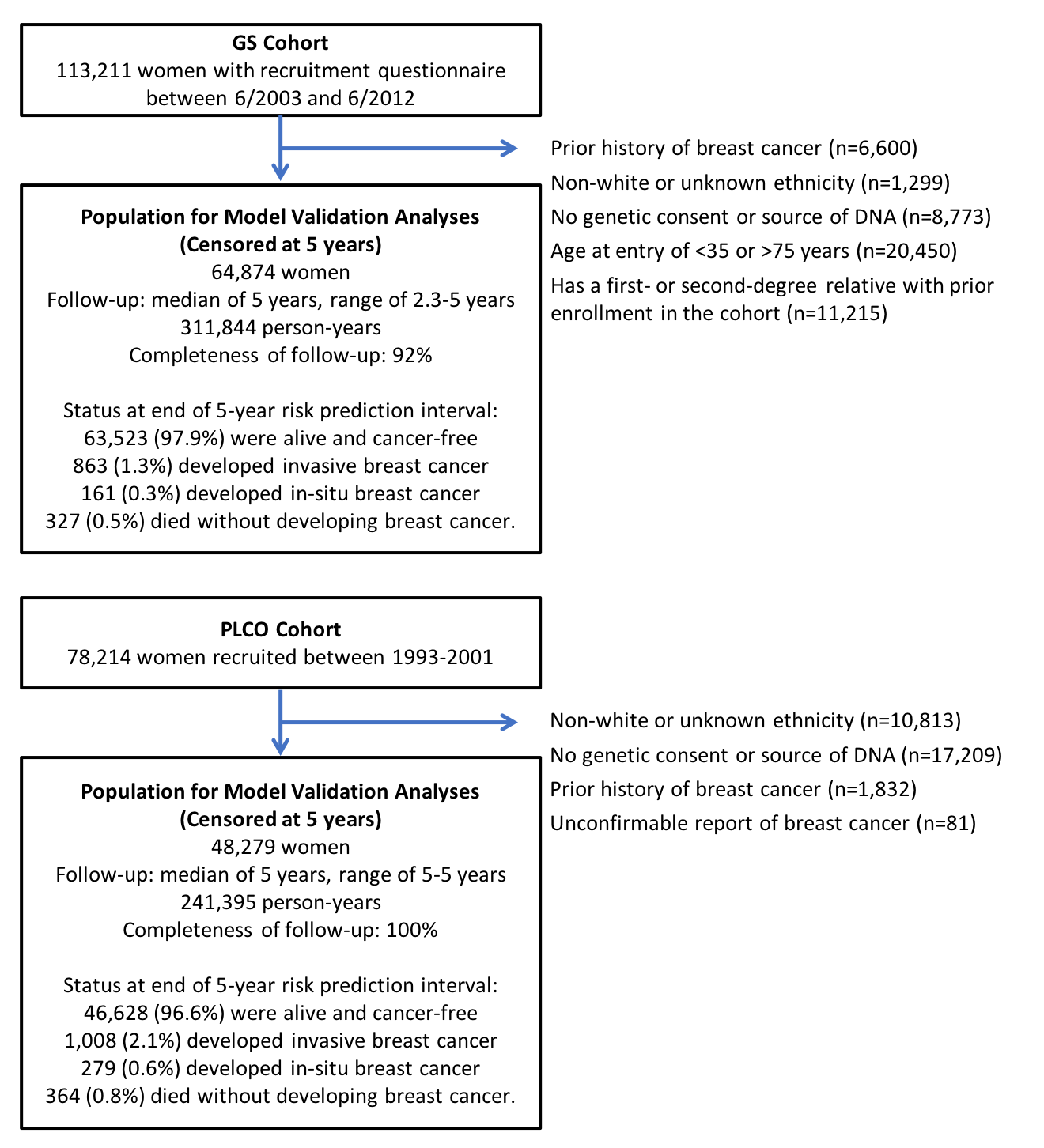
**Supplementary Table 6.** Ratios of expected to observed 5-year absolute risk for the breast cancer risk prediction models in PLCO (1,008 cases, 47,271 non-cases)\*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **AUC (95% CI)** | **Overall** | | | **Top risk decile** | | |
| **O% (95% CI)** | **E (%)** | **E/O ratio (95% CI)** | **O% (95% CI)** | **E (%)** | **E/O ratio (95% CI)** |
| **iCARE-Lit** | 59.8 (58.1 to 61.6) | 2.09 (1.96, 2.22) | 2.46 | 1.18 (1.11 to 1.25) | 3.52 (3.00, 4.04) | 4.72 | 1.34 (1.16 to 1.55) |
| **BCRAT** | 57.2 (55.4 to 59.0) | 1.76 | 0.84 (0.79 to 0.90) | 3.14 (2.65, 3.63) | 3.81 | 1.21 (1.04 to 1.42) |
| **IBIS** | 58.8 (57.1 to 60.6) | 1.47 | 0.70 (0.66 to 0.75) | 3.36 (2.86, 3.87) | 3.02 | 0.90 (0.77 to 1.04) |
| \* AUC = area under the curve, CI = confidence interval, E = expected absolute risk, O = observed absolute risk; PLCO = Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial. The AUCs reported in Supplementary Table 5 are defined based on absolute risk and incorporate the variation due to age. The AUCs (95% CI) based on the relative risk score, which do not include variation of age, are as follows: iCARE-Lit: 58.7 (57.0 to 60.5), BCRAT: 54.9 (53.0 to 56.8), IBIS: 58.1 (56.3 to 59.9). | | | | | | | |

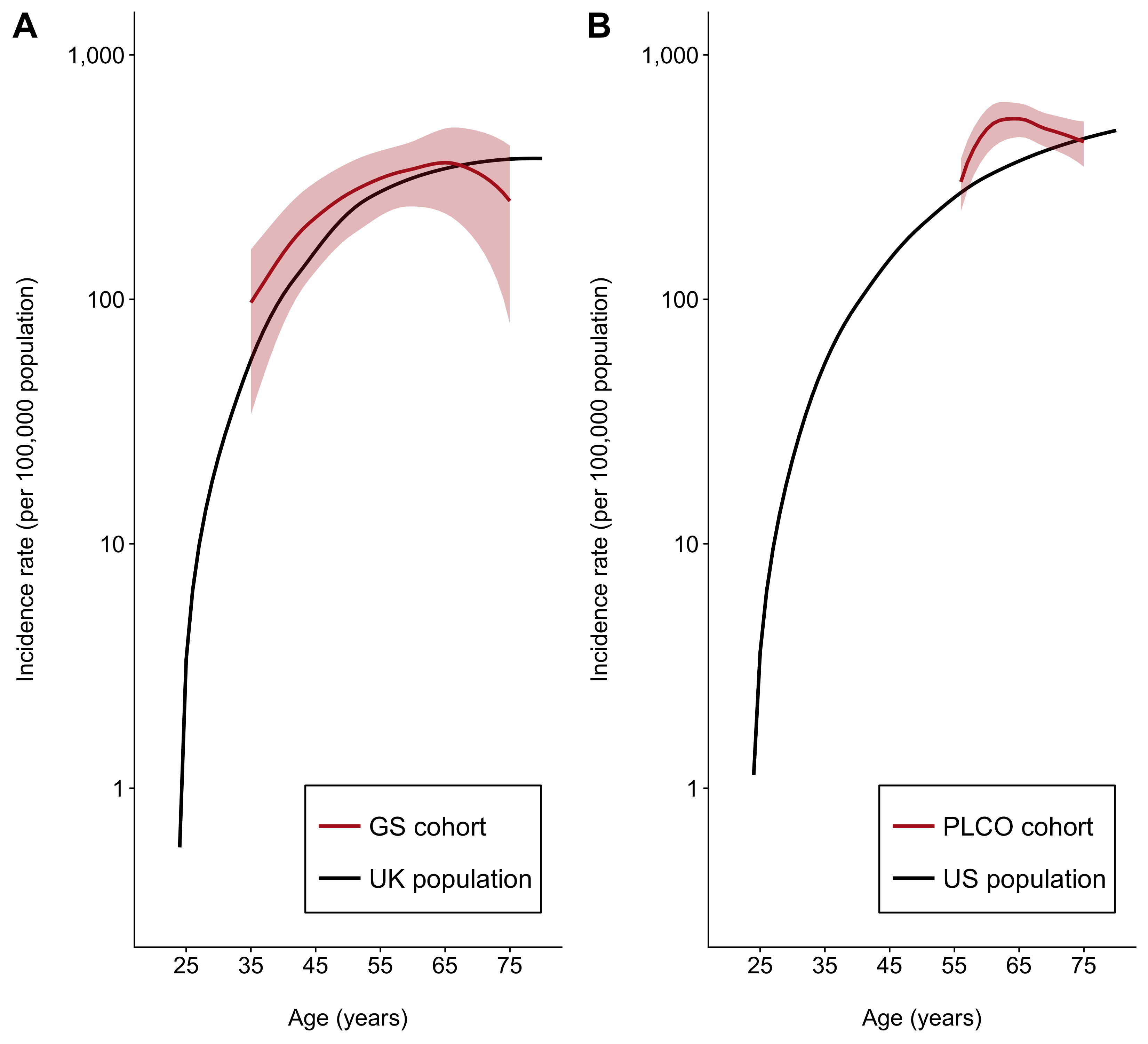
**Supplementary Table 7.** Proportion of at-risk subjects and incident cases expected to be identified at different risk levels for models with different combinations of risk factors, among non-Hispanic White women ages 50-70 in the US\*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Incorporated breast cancer risk factors** | **AUC** | **Very low risk (5-year AR <0.6%)** | | | | **Low risk (5-year AR <1.13%)** | | | | **Moderate to high risk (5-year AR >3%)** | | | | **High risk (5-year AR >6%)** | | | |
| **Total subjects (n)** | | **Cases (n)** | | **Total subjects (n)** | | **Cases (n)** | | **Total subjects (n)** | | **Cases (n)** | | **Total subjects (n)** | | **Cases (n)** | |
| Classical risk factors | 58.5 | 0.1% | (22,548) | <0.1% | (124) | 13.8% | (4,148,896) | 8.2% | (40,516) | 1.7% | (500,167) | 3.4% | (16,819) | <0.1% | (153) | <0.1% | (9) |
| Density | 61.0 | 0.9% | (266,020) | 0.3% | (1,408) | 22.2% | (6,668,229) | 12.3% | (61,062) | 4.3% | (1,304,629) | 9.4% | (46,494) | <0.1% | (7,793) | 0.1% | (519) |
| Current PRS (313-SNPs) | 63.1 | 2.8% | (828,791) | 0.8% | (4,202) | 28.2% | (8,481,751) | 14.7% | (73,379) | 6.8% | (2,030,499) | 15.3% | (76,339) | 0.2% | (45,761) | 0.6% | (3,169) |
| Classical risk factors + Density | 63.7 | 3.6% | (1,074,658) | 1.1% | (5,370) | 29.9% | (8,975,482) | 15.3% | (76,250) | 7.5% | (2,260,220) | 17.3% | (86,518) | 0.2% | (69,272) | 1.0% | (4,862) |
| Classical risk factors + Current PRS | 65.2 | 5.8% | (1,727,019) | 1.7% | (8,349) | 33.4% | (10,041,597) | 16.4% | (81,894) | 9.0% | (2,691,489) | 21.4% | (107,162) | 0.5% | (139,870) | 2.0% | (10,120) |
| Current PRS + Density | 66.7 | 8.6% | (2,579,264) | 2.4% | (11,998) | 36.9% | (11,091,376) | 17.2% | (86,378) | 10.4% | (3,120,884) | 25.9% | (130,076) | 0.8% | (252,651) | 3.8% | (18,955) |
| **All risk factors** | **68.3** | **11.7%** | **(3,513,170)** | **3.1%** | **(15,700)** | **40.0%** | **(12,015,019)** | **17.7%** | **(89,294)** | **11.7%** | **(3,501,245)** | **30.3%** | **(153,112)** | **1.3%** | **(398,463)** | **6.1%** | **(31,083)** |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Improved PRS† | 69.1 | 13.2% | (3,958,089) | 3.4% | (17,356) | 41.3% | (12,403,963) | 17.8% | (90,203) | 12.2% | (3,660,261) | 32.2% | (163,775) | 1.6% | (474,895) | 7.4% | (37,750) |
| Classical risk factors + Improved PRS† | 70.4 | 15.8% | (4,759,774) | 3.9% | (20,169) | 43.5% | (13,057,473) | 17.9% | (91,348) | 13.0% | (3,906,256) | 35.6% | (182,126) | 2.1% | (617,974) | 9.9% | (50,839) |
| Improved PRS + Density†,§ | 71.3 | 17.8% | (5,358,834) | 4.3% | (22,151) | 45.1% | (13,544,458) | 18.0% | (92,127) | 13.4% | (4,021,577) | 37.6% | (192,859) | 2.4% | (712,780) | 11.7% | (60,063) |
| **All risk factors†,§** | **72.2** | **19.9%** | **(5,972,010)** | **4.6%** | **(24,019)** | **46.5%** | **(13,976,097)** | **17.9%** | **(92,231)** | **13.9%** | **(4,182,761)** | **40.2%** | **(207,484)** | **2.8%** | **(834,338)** | **14.0%** | **(72,352)** |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Best PRS‡ | 70.4 | 15.9% | (4,785,777) | 4.0% | (20,268) | 43.6% | (13,096,290) | 17.9% | (91,546) | 12.9% | (3,888,273) | 35.5% | (181,229) | 2.0% | (614,238) | 9.9% | (50,524) |
| Classical risk factors + Best PRS‡ | 71.6 | 18.6% | (5,578,897) | 4.4% | (22,839) | 45.6% | (13,704,261) | 18.0% | (92,219) | 13.6% | (4,077,750) | 38.5% | (197,862) | 2.5% | (754,567) | 12.5% | (64,220) |
| Best PRS + Density‡,§ | 72.7 | 20.8% | (6,261,420) | 4.8% | (24,864) | 47.2% | (14,183,893) | 17.8% | (92,290) | 14.1% | (4,238,237) | 41.2% | (213,450) | 3.0% | (886,501) | 15.0% | (77,885) |
| **All risk factors‡,§** | **73.7** | **23.1%** | **(6,923,329)** | **5.1%** | **(26,697)** | **48.8%** | **(14,655,163)** | **17.7%** | **(92,318)** | **14.4%** | **(4,339,033)** | **43.4%** | **(226,155)** | **3.3%** | **(999,728)** | **17.4%** | **(90,487)** |
| \* The classical risk factors correspond to the iCARE-BPC3 model. Classical risk factors include age at menarche, age at menopause, parity, age at first birth, height, alcohol intake, breast cancer family history, smoking status, BMI, current HRT use, and ever HRT type. The expected number of subjects is calculated using mid-2016 population estimates (n=30,030,821) from the US Census Bureau and the number of cases is calculated using the average predicted five-year risk and the 2015 invasive breast cancer incidence rates from SEER. AUC = area under the curve, AR = absolute risk, PRS = polygenic risk score, SNP = single nucleotide polymorphism. The 0.6% and 1.13% thresholds correspond to the average five-year risk for US women aged 40 years and 50 years, respectively. The 3% threshold is used by the United States Preventive Services Task Force for recommending risk-lowering drugs and 6% is used by the WISDOM trial as a threshold for very high risk.  † PRS including additional SNPs expected to be discovered as current GWAS sample sizes double (i.e., approximately a total of 300,000 cases and 300,000 controls)  ‡ PRS that could explain heritability for breast cancer associated with log-additive effects of all common variants  § Odds ratio for mammographic breast density is adjusted for PRS  The AUC is reported based on the relative risk score in that population and do not incorporate variation of age. | | | | | | | | | | | | | | | | | |

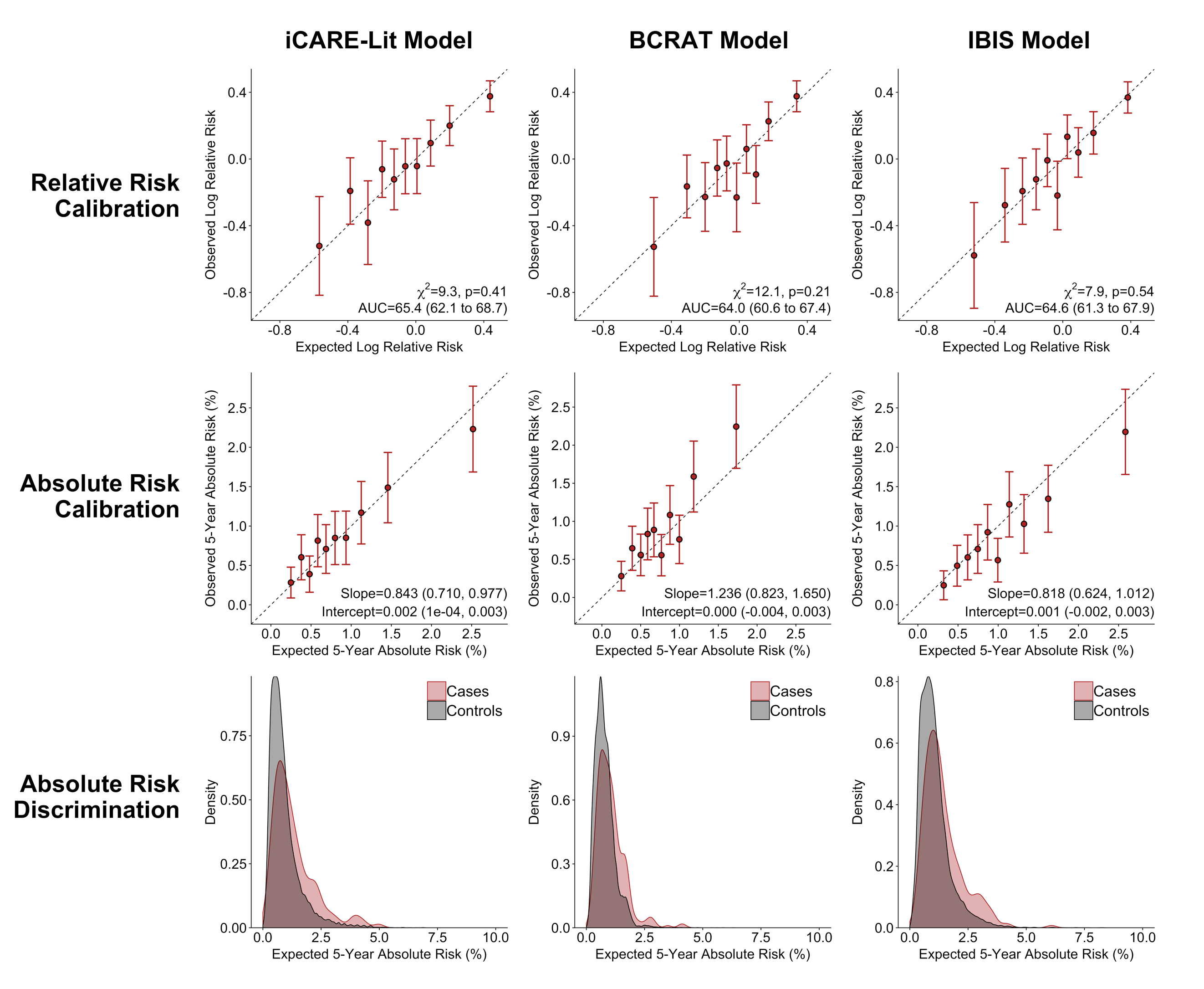
**Supplementary Figure 1.** Study design of validation cohorts

****

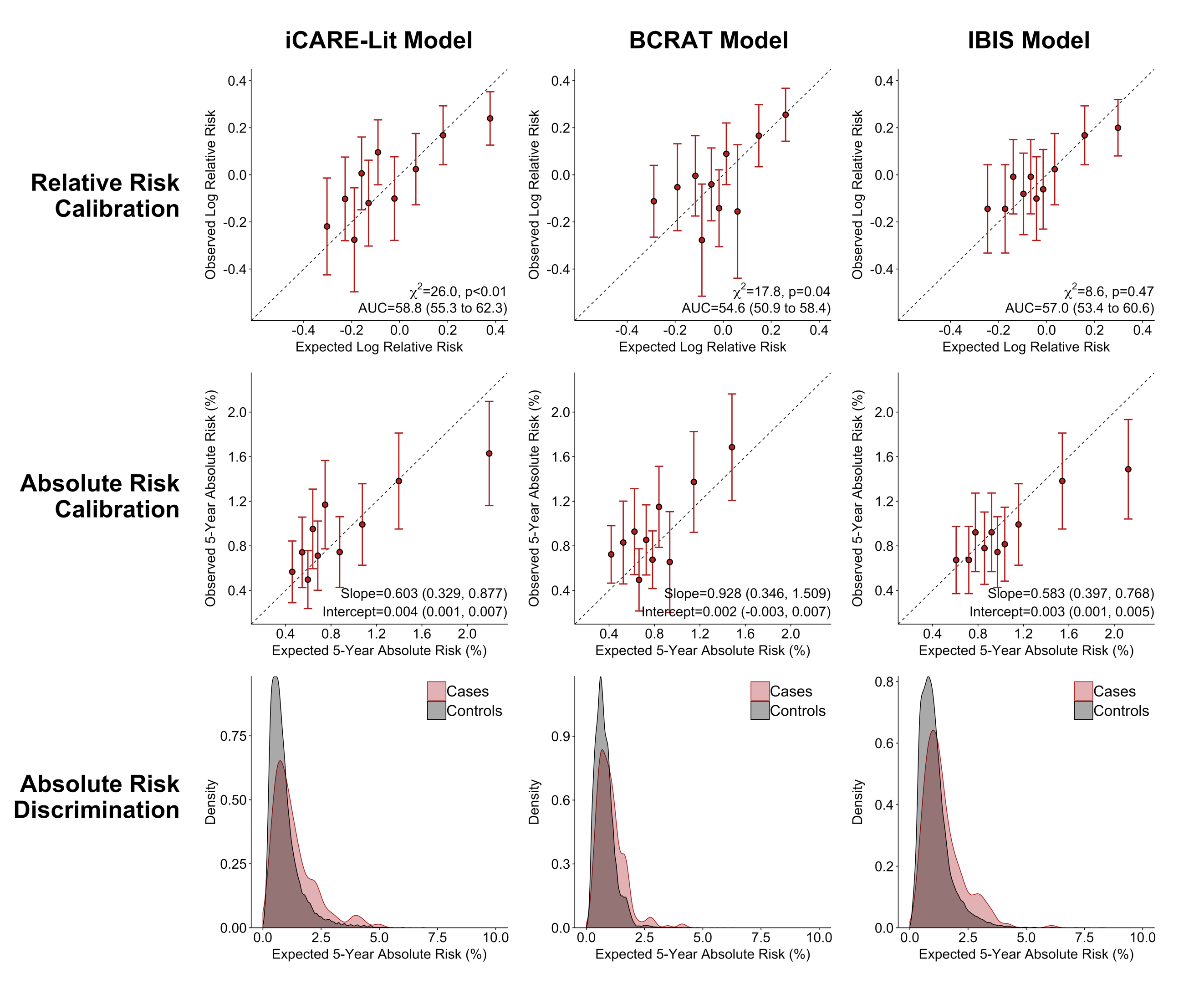
**Supplementary Figure 2.** Breast cancer incidence rates in the validation cohort compared to the general population

****

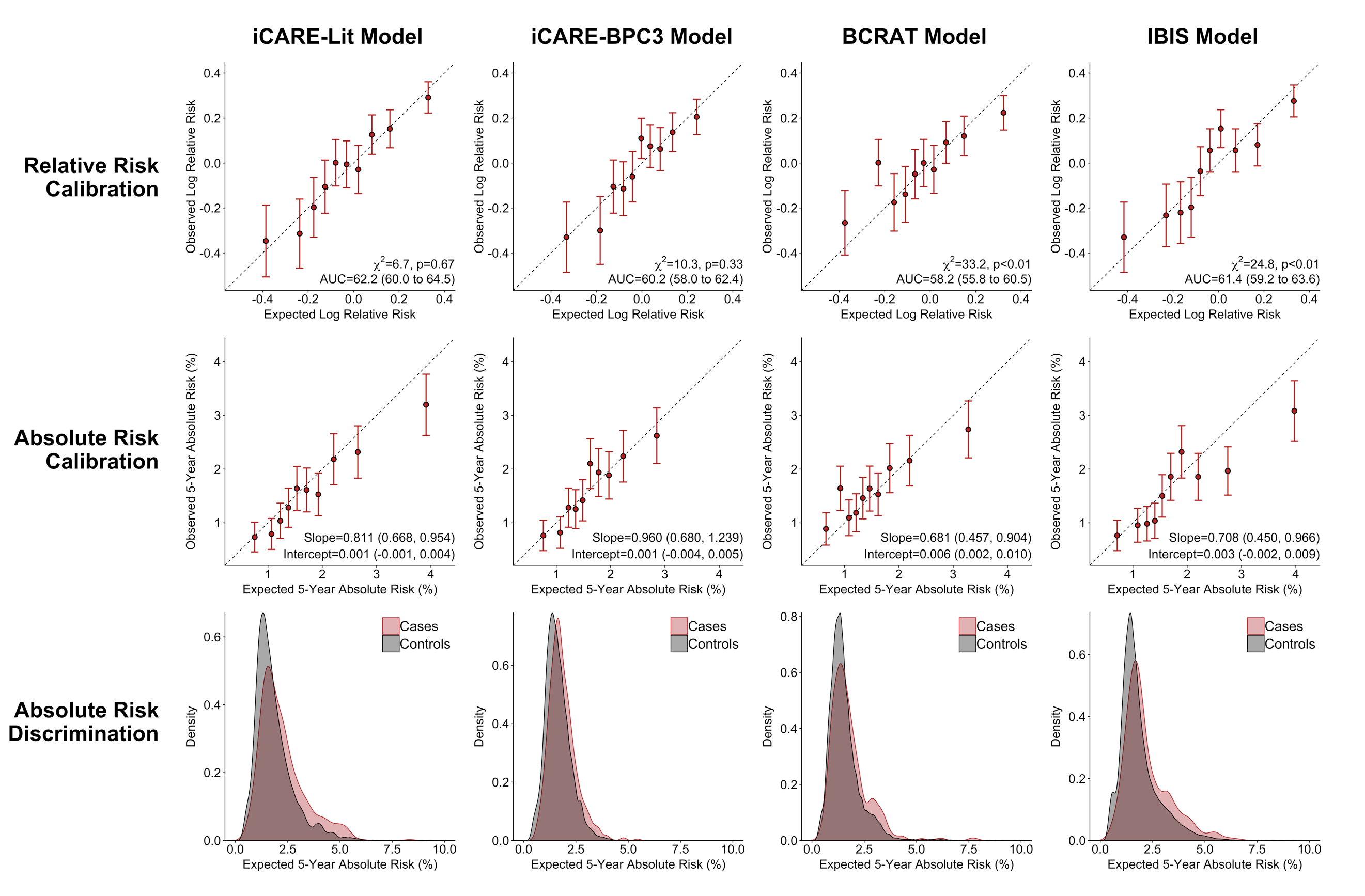
The shaded region represents 95% confidence interval for the breast cancer incidence rates in the validation cohort. **(A)** Breast cancer incidence rates in the Generations Study (GS) cohort compared to the general UK population (ONS, 2006-2010). **(B)** Breast cancer incidence rates in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial cohort compared to the general US population (SEER, 2010-2012).

**Supplementary Figure 3A.** Calibration and discrimination, with risk categories based on deciles of **absolute risk**, of breast cancer risk prediction models in the **GS cohort among women less than 50 years of age**\*

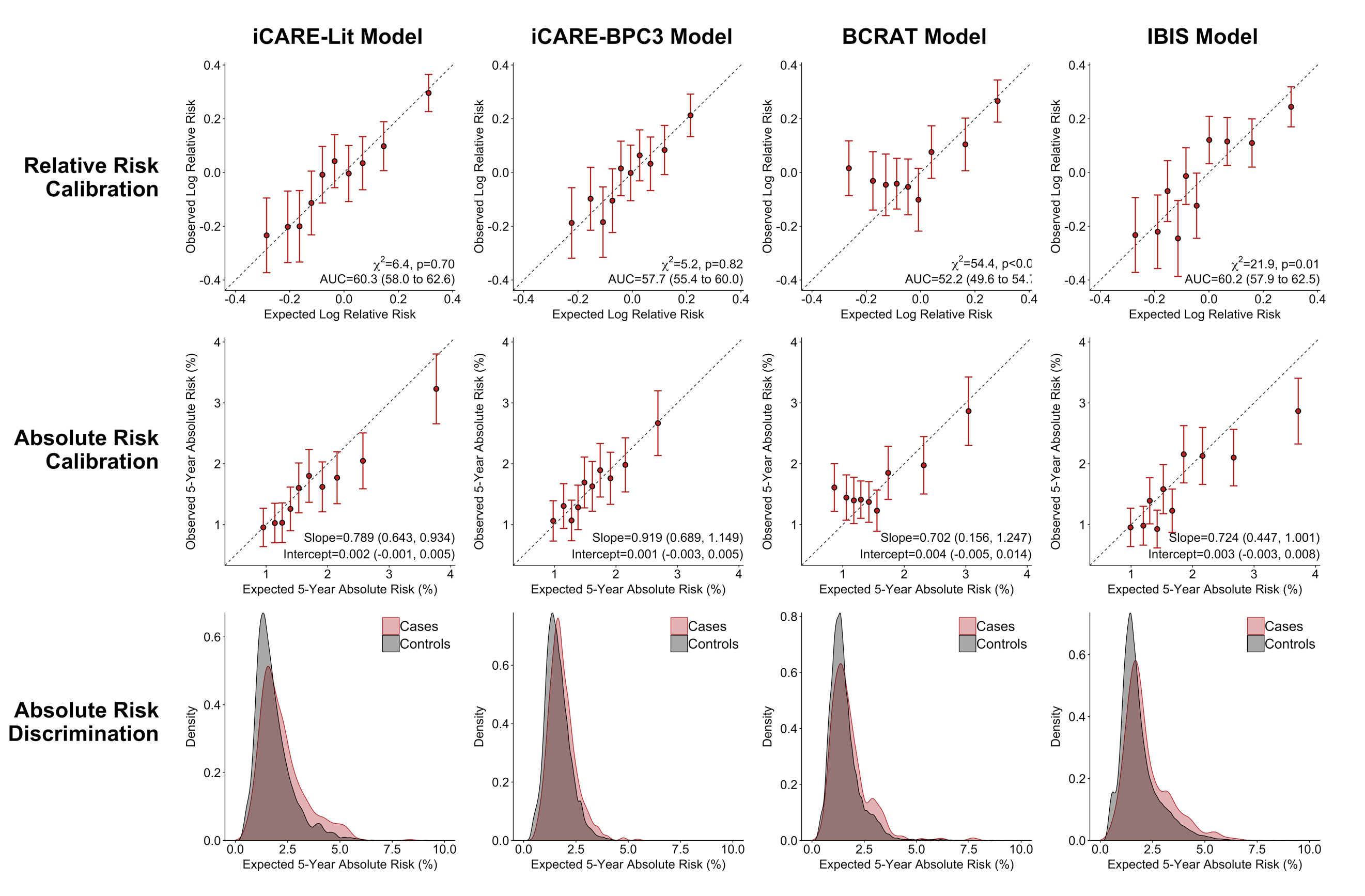
\* The risk categories and AUC are based on absolute risk (Supplementary Figure 3B shows risk categories and AUC based on the relative risk score). AUC = area under the curve, GS = Generations Study, chi-square test statistic.

**Supplementary Figure 3B.** Calibration and discrimination, with risk categories based on deciles of the **relative** **risk score**, of breast cancer risk prediction models in the **GS cohort among women less than 50 years of age**\*

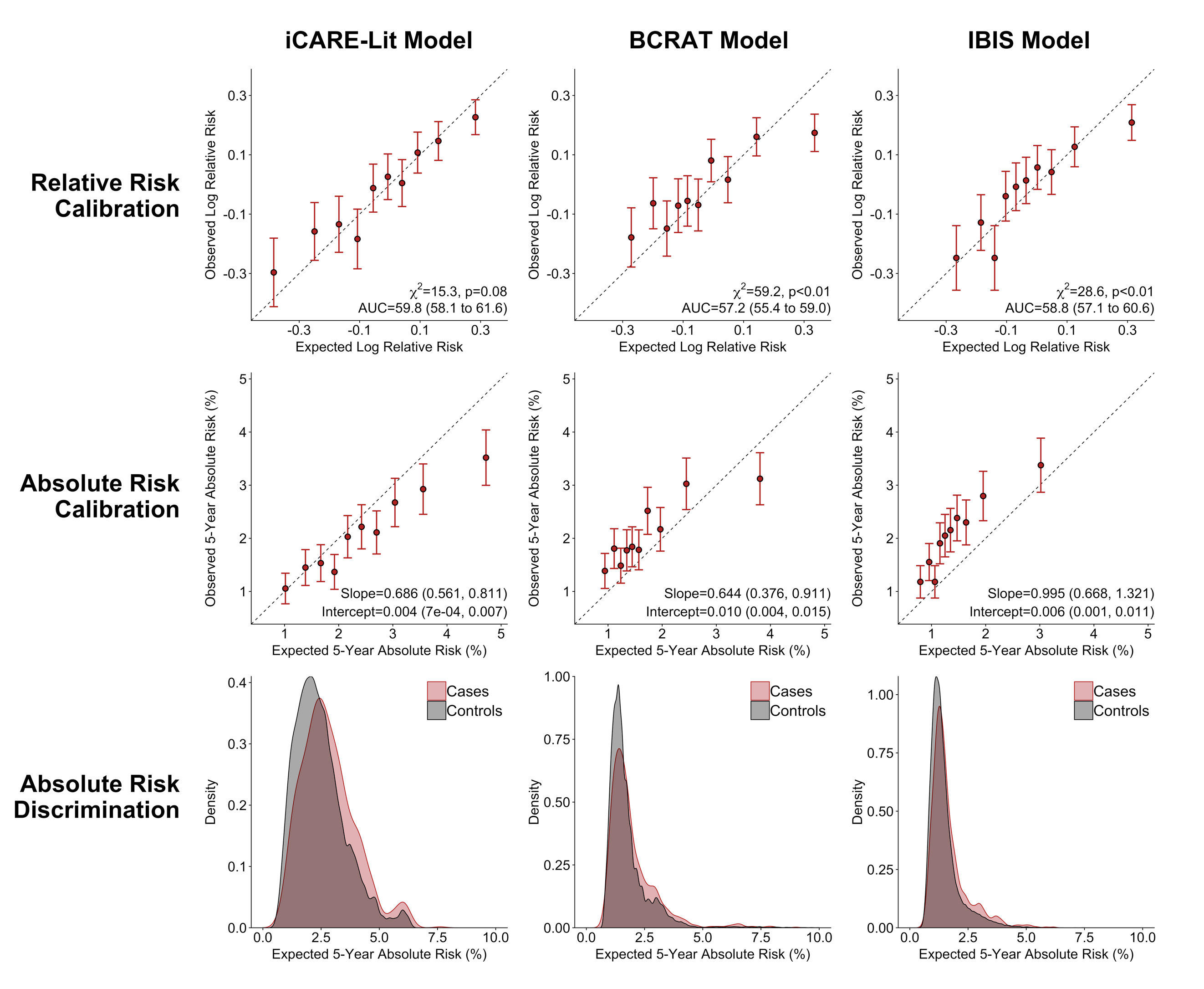
\* The risk categories and AUC are based on the risk score (Supplementary Figure 3A shows risk categories and AUC based on absolute risk). AUC = area under the curve, GS = Generations Study, chi-square test statistic.

**Supplementary Figure 4A.** Calibration and discrimination, with risk categories based on deciles of **absolute risk**, of breast cancer risk prediction models in the **GS cohort among women 50 years of age or greater**\*

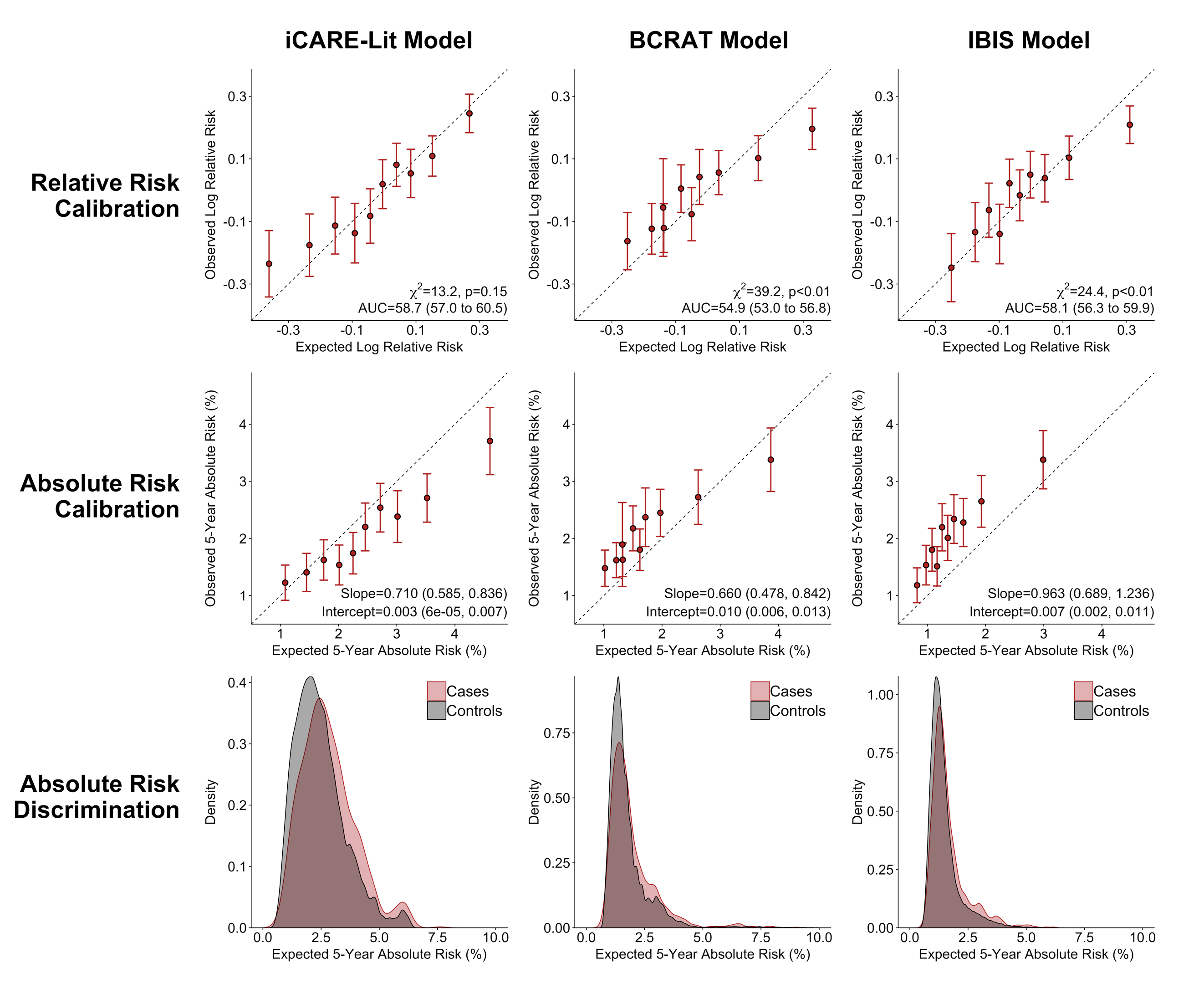
\* The risk categories and AUC are based on absolute risk (Supplementary Figure 4B shows risk categories and AUC based on the relative risk score). AUC = area under the curve, chi-square test statistic.

**Supplementary Figure 4B.** Calibration and discrimination, with risk categories based on deciles of the **relative** **risk score**, of breast cancer risk prediction models in the **GS cohort among women 50 years of age or greater**\*

\* The risk categories and AUC are based on the risk score (Supplementary Figure 4A shows risk categories and AUC based on absolute risk). For the BCRAT model, nine of the ten risk categories are plotted because the fifth risk category included only one case. AUC = area under the curve, chi-square test statistic.

**Supplementary Figure 5A.** Calibration and discrimination, with risk categories based on deciles of **absolute risk**, of breast cancer risk prediction models in the **PLCO cohort among women 50 years of age or greater**\*

\* The risk categories and AUC are based on absolute risk (Supplementary Figure 5B shows risk categories and AUC based on the relative risk score). AUC = area under the curve, chi-square test statistic.

**Supplementary Figure 5B.** Calibration and discrimination, with risk categories based on deciles of the **relative** **risk score**, of breast cancer risk prediction models in the **PLCO cohort among women 50 years of age or greater**\*

\* The risk categories and AUC are based on the risk score (Supplementary Figure 5A shows risk categories and AUC based on absolute risk). AUC = area under the curve, chi-square test statistic.

1. Swerdlow AJ, Jones ME, Schoemaker MJ, et al. The Breakthrough Generations Study: design of a long-term UK cohort study to investigate breast cancer aetiology. *Br J Cancer.* 2011;105(7):911-917.

2. Maas P, Barrdahl M, Joshi AD, et al. Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. *JAMA Oncol.* 2016;2(10):1295-1302.

3. Office for National Statistics. Cancer statistics – Registrations of cancer diagnosed in 2006, England. 2008.

4. Office for National Statistics. Mortality statistics — Deaths registered in 2006, Review of the Registrar General on deaths in England and Wales, 2006. 2008.

5. Surveillance, Epidemiology, and End Results (SEER) Program. SEER\*Stat Database: Incidence - SEER 18 Regs Research Data + Hurricane Katrina Impacted Louisiana Cases, Nov 2015 Sub (2000-2013) <Katrina/Rita Population Adjustment> - Linked To County Attributes - Total U.S., 1969-2014 Counties, National Cancer Institute, DCCPS, Surveillance Research Program. April 2016, based on the November 2015 submission; [www.seer.cancer.gov](file:///Users/garciacm/Google%20Drive/BC%20Modeling%20Validation_Internal/Risk%20Prediction%20Models%20in%20BGS%20Manuscript%20/Draft%20Manuscript/www.seer.cancer.gov).

6. Centers for Disease Control and Prevention, National Center for Health Statistics. Underlying Cause of Death 1999-2016 on CDC WONDER Online Database, released December, 2017. Data are from the Multiple Cause of Death Files, 1999-2016, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program. Accessed at <http://wonder.cdc.gov/ucd-icd10.html> on Feb 27, 2018.

7. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst.* 1989;81(24):1879-1886.

8. Costantino JP, Gail MH, Pee D, et al. Validation studies for models projecting the risk of invasive and total breast cancer incidence. *J Natl Cancer Inst.* 1999;91(18):1541-1548.

9. Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. *Stat Med.* 2004;23(7):1111-1130.

10. Pal Choudhury P, Maas P, Wilcox A, et al. iCARE: An R Package to Build, Validate and Apply Absolute Risk Models. *bioRxiv.* 2018. <https://doi.org/10.1101/079954>.

11. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988;44(3):837-845.

12. Surveillance, Epidemiology, and End Results (SEER) Program ([www.seer.cancer.gov](file:///Users/garciacm/Google%20Drive/BC%20Modeling%20Validation_Internal/Risk%20Prediction%20Models%20in%20BGS%20Manuscript%20/Draft%20Manuscript/www.seer.cancer.gov)) SEER\*Stat Database: Incidence - SEER 18 Regs Research Data + Hurricane Katrina Impacted Louisiana Cases, Nov 2017 Sub (2000-2015) <Katrina/Rita Population Adjustment> - Linked To County Attributes - Total U.S., 1969-2016 Counties, National Cancer Institute, DCCPS, Surveillance Research Program. April 2018, based on the November 2017 submission.

13. Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BA. Polygenic susceptibility to breast cancer and implications for prevention. *Nat Genet.* 2002;31(1):33-36.

14. Chatterjee N, Wheeler B, Sampson J, Hartge P, Chanock SJ, Park JH. Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. *Nat Genet.* 2013;45(4):400-405, 405e401-403.

15. Chatterjee N, Shi J, Garcia-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat Rev Genet.* 2016;17(7):392-406.

16. Final Recommendation Statement: Breast Cancer: Medications for Risk Reduction. U.S. Preventive Services Task Force. 2016; <https://www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/breast-cancer-medications-for-risk-reduction>. Accessed December 13, 2018.

17. Shieh Y, Eklund M, Madlensky L, et al. Breast Cancer Screening in the Precision Medicine Era: Risk-Based Screening in a Population-Based Trial. *J Natl Cancer Inst.* 2017;109(5).

18. Practice Bulletin Number 179: Breast Cancer Risk Assessment and Screening in Average-Risk Women. *Obstet Gynecol.* 2017;130(1):e1-e16.

19. NCCN Clinical Practice Guidelines in Oncology: Breast Cancer Screening and Diagnosis, Version 1.2018. In: National Comprehensive Cancer Network ([www.nccn.org](file:///Users/garciacm/Google%20Drive/BC%20Modeling%20Validation_Internal/Risk%20Prediction%20Models%20in%20BGS%20Manuscript%20/Draft%20Manuscript/www.nccn.org)); 2018.

20. Oeffinger KC, Fontham ET, Etzioni R, et al. Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update From the American Cancer Society. *Jama.* 2015;314(15):1599-1614.

21. Michailidou K, Lindstrom S, Dennis J, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature.* 2017;551(7678):92-94.

22. Garcia-Closas M, Gunsoy NB, Chatterjee N. Combined associations of genetic and environmental risk factors: implications for prevention of breast cancer. *J Natl Cancer Inst.* 2014;106(11).

23. Milne RL, Kuchenbaecker KB, Michailidou K, et al. Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. *Nat Genet.* 2017;49(12):1767-1778.

24. Mavaddat N, Michailidou K, Dennis J, et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am J Hum Genet.* 2018.

25. Zhang Y, Qi G, Park JH, Chatterjee N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nat Genet.* 2018;50(9):1318-1326.

26. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology.* 2010;21(1):128-138.

27. Rousson V, Zumbrunn T. Decision curve analysis revisited: overall net benefit, relationships to ROC curve analysis, and application to case-control studies. *BMC Med Inform Decis Mak.* 2011;11:45.

28. Zhang Z, Rousson V, Lee WC, et al. Decision curve analysis: a technical note. *Ann Transl Med.* 2018;6(15):308.

29. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Medical Decision Making.* 2006;26(6):565-574.

30. Centers for Disease Control and Prevention, US Department of Health and Human Services. 2010 National Health Interview Survey (NHIS) Public Use Data Release, NHIS Survey Description. 2011. <https://www.cdc.gov/nchs/nhis/data-questionnaires-documentation.htm>.

31. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. *Lancet Oncol.* 2012;13(11):1141-1151.

32. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. *Control Clin Trials.* 1998;19(1):61-109.

33. Cohort Fertility Tables in England and Wales, 2011. Office for National Statistics. <http://data.gov.uk/dataset/cohort_fertility_england_and_wales> Accessed December 22, 2015.

34. Reeves GK, Pirie K, Green J, Bull D, Beral V. Comparison of the effects of genetic and environmental risk factors on in situ and invasive ductal breast cancer. *Int J Cancer.* 2012;131(4):930-937.

35. Centers for Disease Control and Prevention (CDC). National Health and Nutrition Examination Survey Questionnaire. 2008-2012. <https://wwwn.cdc.gov/nchs/nhanes/Default.aspx>.

36. National Centre for Social Research, University College London. Department of Epidemiology and Public Health, 2011, Health Survey for England, 2006, [data collection], UK Data Service, 4th Edition, Accessed December 15, 2015. SN: 5809, <http://doi.org/10.5255/UKDA-SN-5809-1>.

37. National Centre for Social Research, University College London. Department of Epidemiology and Public Health, 2011, Health Survey for England, 2005, [data collection], UK Data Service, 3rd Edition, Accessed March 2, 2016. SN: 5675, <http://doi.org/10.5255/UKDA-SN-5675-1>.

38. Hunter DJ, Colditz GA, Hankinson SE, et al. Oral contraceptive use and breast cancer: a prospective study of young women. *Cancer Epidemiol Biomarkers Prev.* 2010;19(10):2496-2502.

39. Nelson HD, Zakher B, Cantor A, et al. Risk factors for breast cancer for women aged 40 to 49 years: a systematic review and meta-analysis. *Annals of internal medicine.* 2012;156(9):635-648.

40. Green J, Cairns BJ, Casabonne D, Wright FL, Reeves G, Beral V. Height and cancer incidence in the Million Women Study: prospective cohort, and meta-analysis of prospective studies of height and total cancer risk. *Lancet Oncol.* 2011;12(8):785-794.

41. National Centre for Social Research, University College London. Department of Epidemiology and Public Health, 2010, Health Survey for England, 2002, [data collection], UK Data Service, 2nd Edition, Accessed March 2, 2016. SN: 4912, <http://doi.org/10.5255/UKDA-SN-4912-1>.

42. Hamajima N, Hirose K, Tajima K, et al. Alcohol, tobacco and breast cancer--collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease. *Br J Cancer.* 2002;87(11):1234-1245.

43. Tice JA, Miglioretti DL, Li CS, Vachon CM, Gard CC, Kerlikowske K. Breast Density and Benign Breast Disease: Risk Assessment to Identify Women at High Risk of Breast Cancer. *J Clin Oncol.* 2015;33(28):3137-3143.

44. Hartmann LC, Sellers TA, Frost MH, et al. Benign breast disease and the risk of breast cancer. *N Engl J Med.* 2005;353(3):229-237.

45. Pharoah PD, Lipscombe JM, Redman KL, Day NE, Easton DF, Ponder BA. Familial predisposition to breast cancer in a British population: implications for prevention. *Eur J Cancer.* 2000;36(6):773-779.

46. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *Lancet.* 2001;358(9291):1389-1399.

47. Parkin DM. Is the recent fall in incidence of post-menopausal breast cancer in UK related to changes in use of hormone replacement therapy? *Eur J Cancer.* 2009;45(9):1649-1653.

48. Beral V, Reeves G, Bull D, Green J. Breast cancer risk in relation to the interval between menopause and starting hormone therapy. *J Natl Cancer Inst.* 2011;103(4):296-305.

49. Opatrny L, Dell'Aniello S, Assouline S, Suissa S. Hormone replacement therapy use and variations in the risk of breast cancer. *Bjog.* 2008;115(2):169-175; discussion 175.